

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

A work of stiction

The relationship between static friction and adhesion modelled in a single drop of liquid **PAGE 676**

NUCLEAR TRANSFER

THE DAY THEY MADE DOLLY

The story of the first adult clone, twenty years on

PAGE 604

POLICYMAKING

VISIONS OF UTOPIA

Can another international panel help social progress?

PAGE 616

MODERN WARFARE

MARCH OF THE DRONES

The destruction of the accepted rules of war

PAGE 618

NATURE.COM/NATURE

30 June 2016 £10

Vol. 534, No. 7609



THIS WEEK

EDITORIALS

GENE THERAPY Past tragedy continues to haunt trial prospects **p.590**

WORLD VIEW Reform of Indian education must address plagiarism **p.591**



AGEING Old monkeys grow less interested in new tricks **p.593**

Non-expert nation

Scientists — just like everybody else — have little idea what will happen now that the United Kingdom has voted to exit the European Union.

Psychologists who have studied the peculiar phenomenon of buyer regret — the second thoughts that follow the purchase of a shiny new car, say — note a curious paradox. The more effort that consumers put into making their decision, the more information they seek and the more they weigh up the options, the more likely they are to want to change their mind later.

Just how much careful thought the people of the United Kingdom put into last week's decision to quit the European Union is currently a matter of some debate. But if the prominent examples of buyer regret among people who voted 'Leave' and now want to 'Remain' are any guide, it may have been more than many critics think.

Psychologists might conclude that Kelvin Mackenzie, the former editor and now columnist of *The Sun* newspaper, must have been weighing up the options very carefully indeed when he wrote his "10 reasons why you must vote Brexit" the week before the crucial vote. How else to explain his U-turn, a few days after 52% of voters heeded his demand, when he admitted: "I have buyer's remorse. A sense of being careful what you wish for. To be truthful I am fearful of what lies ahead."

Scientists in the United Kingdom and elsewhere share his anxiety — and fear. Hundreds have responded to calls from this journal to express their feelings, and the overwhelming question that they have replied with is: what happens now?

UK politicians who pushed for the country to exit the EU have gone to ground. A similar silence reigns in the European Commission's research directorate. Commission sources mutter darkly, and only off the record, of 'uncharted territories' and 'needing time' to consider the many issues that will arise. UK politicians and the research directorate declined to engage before the vote with the 'what if' question, at least publicly. So it is no surprise that scientists have been left with the feeling that no one had planned for the Brexit eventuality. What will be the status of those from other EU countries doing their PhDs or postdoctoral research in the United Kingdom? What will happen to the EU-funded research collaborations that are led from the United Kingdom?

What do we know for sure? Some of the most familiar European research facilities are not creatures of the EU, so will remain fundamentally unaffected by Brexit. These include the European particle-physics laboratory CERN, the European Molecular Biology Laboratory and the European Space Agency.

More recently, the European Commission has found a way to steer the creation of other, much-needed Europe-wide research infrastructures through an umbrella structure called ESFRI (European Strategy Forum on Research Infrastructures) that helps to foster intergovernmental agreements in which it has no fundamental role.

Some research infrastructures are based on a particular legal framework that stipulates that the host country must be a member state. For the European Spallation Source, headquartered in Sweden, and the Biobanking and BioMolecular resources Research Infrastructure headquartered in Austria, nothing changes. For the European Social

Survey and the structural-biology infrastructure known as Instruct, both headquartered in the United Kingdom, Brexit means that new arrangements will have to be made; internal talks have already begun.

Talks on similar agreements for core European Commission scientific activities won't start until the United Kingdom formally declares its exit by triggering the much-discussed article 50 of the EU treaty. When (and if) that will happen depends on how quickly the country resolves various questions of its own: not least, who the next prime minister will be, the proper legal route, and the broader constitutional question of whether it should follow through on a democratic decision that seems likely to damage the prospects of so many who voted for it.

If the United Kingdom does trigger article 50, research facilities owned by the commission and stationed in the country, such as the nuclear-fusion facility JET, face an uncertain future. And until a new agreement is made, UK scientists will be shut out of the EU's multi-billion-euro Horizon 2020 programme — including its prestigious European Research Council granting body, from which the United Kingdom benefits more than any other country, by a wide margin.

Michael Gove, a senior figure in the Leave camp, notoriously claimed during the campaign that the United Kingdom has "had enough of experts". He has got his wish, but he should beware: buyer regret is not available to those who did the selling. ■ **SEE NEWS P.597**

The big picture

Interdisciplinary research is vital if we are to meet the diverse needs of modern society.

To tackle society's challenges through research requires the engagement of multiple disciplines. For two examples, in responding to the challenges of climate change and of social progress, see the Comment articles on pages 613 and 616, respectively.

To highlight the issues that arise in such research, imagine an integrated project to determine the causes of destructive risk-taking in inner-city adolescents and to identify appropriate interventions. Such a programme might combine disciplines ranging from anthropology, sociology, psychology, law, economics and ethics to psychiatry, health systems, urban design and developmental neurobiology.

To frame the research challenge, and to design interventions that will be effective in targeted neighbourhoods, academic researchers need to work with non-academic partners to understand the needs

of the community, the political context and the barriers — structural and behavioural — to applying the lessons that might be learned. The researchers would also need to learn how colleagues from other disciplines approach the issues and frame the research questions in a mutually acceptable way. They must also learn to respect what is possible in each discipline, and how insights are gained and possible implementations are made. All this is easier said than done, but it is essential.

Funders must rise to the challenge of supporting these tough research necessities. That means having enough of an overview of a project to oversee the selection of peer reviewers whose individual perspectives will inevitably be narrower than those of the project. An ideal funder would also include potential users of the project's outcome among its assessors, to ensure that the research has practical impact as well as academic weight.

The world is ill-equipped to uphold such ideals. For example, a paper published in this issue of *Nature* (R. Bromham *et al.* *Nature* 534, 684–687; 2016) provides evidence that multidisciplinary research is less attractive to funders than single-discipline research. The work is based on an analysis of grant applications to the Australian Research Council, but there is every reason to believe that the conclusion can be generalized. The metrics of interdisciplinarity introduced by the authors can also serve as warning indicators for funders, telling them when they need to take special measures to do a project justice.

The good news is that many funding agencies are aware of the challenge, and of how far they need to go to meet it. The Global Research Council (GRC) is a forum in which government funders discuss their common challenges. At its annual meeting in Delhi last month, the focus was on interdisciplinarity. The council commissioned a survey and analysis of the practices of many funders. It also issued a

statement of principles on interdisciplinarity (go.nature.com/290mqqt).

The GRC is not a decision-making body. But it was evident at the meeting that the funders recognize the need for new measures. An obvious one is that grants should last long enough for interdisciplinary research to take shape. Another is that funding agencies should have a good enough grasp of the subject matter to ensure that a well-informed, multidisciplinary assessment can be conducted.

Journals, too, must face up to such challenges. *Nature* and its research journals take pride in their capacity to handle interdisciplinary research. The multidisciplinary editorial teams see it as part of their job to do so — in selecting referees from diverse disciplines, and in considering their comments within the framing of the paper under discussion, rather than that of the individual assessors. In such a context, it is not unknown for *Nature's* editors to overrule all referees' recommendations against publication of a technically valid paper, and to publish it.

What is more, the *Nature* journals are recruiting social scientists to address our editorial goal of increasing the attention given to the societal challenges of sustainability and health. *Nature* itself will soon be recruiting social-sciences editors. In launching *Nature Climate Change* and *Nature Energy*, and as we recruit for the launch of *Nature Human Behaviour* next year, we have already learned some important lessons about the sense of professional identity of sociologists, anthropologists, economists and psychologists.

Without that developing sense of respect for diverse types of quantitative and qualitative research, progress by funders, publishers and universities in interdisciplinary research will founder. ■

“The good news is that many funding agencies are aware of the challenge.”

Calculated risks

Gene-therapy trials must move forward, but not without due consideration of the dangers.

Jesse Gelsinger was 18 and healthy when he died in 1999 during a gene-therapy experiment. He had a condition called ornithine transcarbamylase deficiency (OTC), but it was under control through a combination of diet and medication. Like others with the disorder, Gelsinger lacked a functional enzyme involved in breaking down ammonia, a waste product of protein metabolism that becomes toxic when its levels become too high. The gene therapy that he received used a viral vector to introduce a normal gene for the enzyme.

Gene therapy remains an obvious route to treat OTC. Simply adding the missing gene has been shown to repair metabolism in mice. But the memory of what happened to Gelsinger has slowed progress in gene therapy for any condition.

That memory was firmly on the agenda at a meeting of the US National Institutes of Health's Recombinant DNA Advisory Committee (RAC) last week. The RAC evaluates proposals to use modified DNA in human trials, and presenting to it were Cary Harding, a medical geneticist at Oregon Health and Science University in Portland, and Sam Wadsworth, chief scientific officer at Dimension Therapeutics in Cambridge, Massachusetts. The duo were proposing the first new trial of gene therapy for OTC.

Harding and the researchers at Dimension argue that the technology and our understanding of physiology have advanced enough since 1999 to try it again in people. Gelsinger died after his body overreacted to the vector used to introduce the OTC gene. Dimension's therapy uses a different viral vector, called AAV8, which has been tested numerous times in people with other conditions, with few adverse effects.

Such assurances were not enough for the RAC, and particularly not

for its bioethicists and historians. Dawn Wooley, a virologist at Wright State University in Dayton, Ohio, pointed out that an RAC panel raised concerns about Gelsinger's trial in 1995, but decided to let the test go ahead. “We can't let it happen again, we cannot,” she says.

Perhaps the greatest indication of how Gelsinger's death haunts the RAC came when one member suggested that the researchers explain in the consent form to be sent to prospective participants that someone had died in a similar study and attracted media attention.

There are some scientific reasons to be careful. AAV8 can cause mild liver toxicity in healthy people, and the steroids used to treat that could lead to complications in people with OTC. With so little known about these effects, the RAC members suggested that the researchers lower the dose to one that is more likely to be safe, even if it is potentially not effective.

After some discussion, the RAC voted unanimously to approve the trial. However, that came with a long list of conditions, including that the treatment first be tested in a second animal species. The researchers disagree with most of the conditions, believing that more expensive animal trials will add nothing. They feel that they are being held to a different standard from most trials.

Dimension still plans to submit an application to the US Food and Drug Administration (FDA) later this year to start a clinical trial. It is unclear how heavily the RAC's recommendations weigh into FDA decisions, but Wadsworth says that the company will conduct its trials overseas if necessary. “These patients have been waiting a long time,” he says.

He is right. Therapies can be tested in non-human animals only for so long — at some point, volunteers such as Gelsinger must step forward. Yet the echoes of a trial done 17 years ago cannot be easily silenced. In fact, Gelsinger's name came up several times at the RAC meeting. Researchers from the University of Pennsylvania in Philadelphia had even mentioned him earlier that morning, when proposing the first human trial of CRISPR gene-editing technology as a treatment for cancer. The RAC approved that proposal, but its implication was clear: take care. Avoidable failures could stymie CRISPR research for decades. History must not repeat itself. ■



Stop teaching Indians to copy and paste

Major reform of education in India should encourage original thinking to boost the nation's research, argues **Anurag Chaurasia**.

My eight-year-old son came home from school disappointed last week. When asked the test question “How can we save the environment from pollution?”, he had tried to write the answer in his own way. This did not go down well with his teacher, who cut his mark and asked why he had not repeated the answer as it was printed in the textbook. That's common practice in India. To get top marks, school children must learn and regurgitate answers presented to them. With such a culture, is it any wonder that plagiarism and unoriginal thinking are so prevalent in Indian science and research?

We should all be disappointed with my son's experience at his school. And India currently has a rare, possibly once-in-a-lifetime opportunity to sort it out. A major review of the nation's education system has made several recommendations to the government, which has so far not published them. Scientists and others are now waiting for the government to say what it will do.

The education system is the best place to start to improve Indian science. Many of the problems that hold back Indian research are set in motion when researchers are in school and university. Science, they are told and shown, is about answering questions, not asking them. Even at university level, we are taught to learn from the class notes written by the teachers on the board, who themselves copy it from a book, and to answer in the same way in the examination.

This slack attitude goes right to the top. Successful Indian grant applications often copy text from grants submitted in other countries. And a 2010 report on genetically modified crops prepared by officials from six Indian science academies simply cut-and-pasted text from a previous publication. India doesn't take the offence seriously. Researchers who are shown to have committed plagiarism — which is serious misconduct, and enough in many countries to end a career — are typically given only a note of instruction not to do it again.

India had a rich education system in the past, which gave the world many influential thinkers and writers. The coming school reform must attempt to reinstate the once-prized qualities of innovation and discovery. Perhaps this change will also help to kick-start interest in the fundamental sciences, which have become less popular in recent years as students switch to applied sciences, medicine and commerce.

Higher-education institutions can make changes that will have a more immediate impact on Indian science. They must take a harder line on plagiarism by setting and enforcing rules and by introducing ethics classes to show students that the practices that they learned in school are no longer acceptable. And the government must demand that universities introduce more and stricter measures to guarantee the standard of the degrees, and especially the postgraduate qualifications they issue.

Poor standards explain why Indian universities rarely feature, and sometimes aren't included at all, in league tables of international institutions. (The 2015–16 Times Higher Education World University Rankings do not include a single Indian university in the top 200.) This is unacceptable for a nation of India's size and ambition.

Some of the best institutions in the country have taken a few steps to improve quality — they insist that a PhD project must produce two papers in international journals, and that a thesis is reviewed by a foreign expert. But these measures are too easily circumvented. PhD students simply pay to publish in a low-quality, open-access journal and send their thesis to a friend. We need stricter definitions of who can publish and review work that will grant a young Indian scientist a ticket to academia.

This is especially true for the private education institutions and publishers that are rapidly emerging, and that are polluting Indian science and scientific literature just to make money. These institutions charge students to complete low-grade PhDs and to publish poor work — a move encouraged by government officials who want to give private education more autonomy. Moves to allow foreign institutions to establish campuses in India must be closely regulated if they are not to make the situation worse.

There are at least some welcome attempts under way to improve journal quality. The University Grants Commission has asked experts to produce a list of approved journals in which academics must publish to earn points in the Indian system that is used to judge performance and award promotions. This idea should be extended

to include papers that are published as part of a PhD programme.

The final change that the education reform can bring about for Indian science is to alter the selection and attitudes of scientists who make it to tenured positions. At present, too many see science as a route to a stable career in administration. They want to leave the laboratory at the earliest possible opportunity — perhaps because they have never learned the true nature and satisfaction of a research job well done. In my 20-year scientific career, I have rarely seen any researchers who wish to work in the lab instead of opting for a desk job. Most of the best Indian scientists initially did very well at the bench but soon went into administration, losing their talent in the office files.

I don't know whether my son will want to be a scientist. But if he does, I want him to be a true scientist — and in India, that will demand big changes in the way that he is taught. ■

Anurag Chaurasia is a biotechnologist with the National Bureau of Agriculturally Important Microorganisms in Kushmaur, India.
e-mail: anurag_vns1@yahoo.co.in

MANY OF THE PROBLEMS THAT HOLD BACK INDIAN RESEARCH ARE SET IN MOTION WHEN RESEARCHERS ARE IN SCHOOL AND UNIVERSITY.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PLANT BIOTECHNOLOGY

Tobacco plants make malaria drug

Inexpensive fast-growing plants have been transformed into factories that churn out an important antimalarial drug.

Artemisinin is the only proven malaria treatment, with hundreds of millions of doses taken every year. The sweet wormwood plant (*Artemisia annua*) produces a precursor of the compound, artemisinic acid, only in low quantities, and is expensive to grow. To scale-up production, a team led by Ralph Bock at the Max Planck Institute of Molecular Plant Physiology in Potsdam-Golm, Germany, inserted genes for artemisinic acid synthesis into tobacco plants' chloroplasts — abundant organelles that have their own DNA. By adding 'accessory genes' that make artemisinic acid production more efficient, they created a line that pumps out 120 milligrams of artemisinic acid per kilogram of biomass.

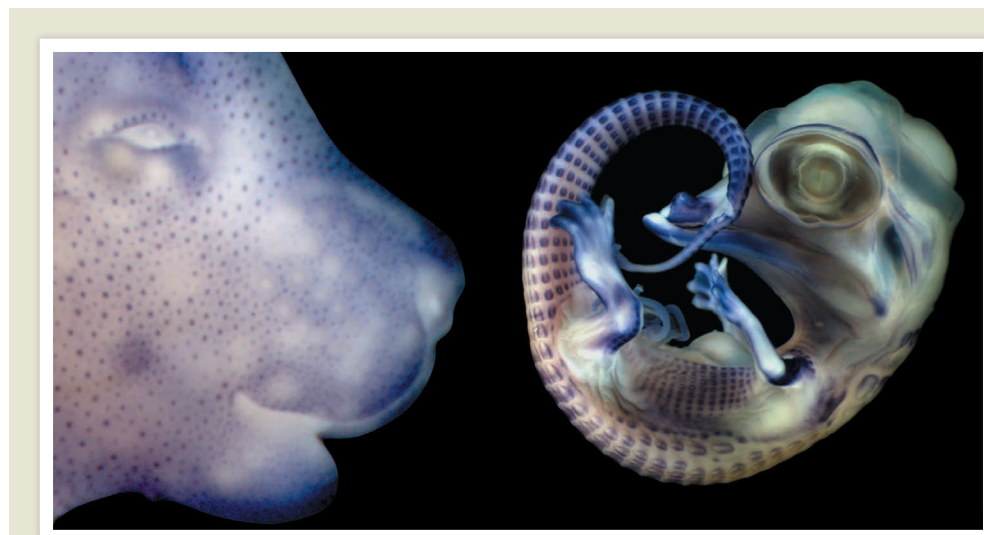
The researchers estimate that the world's demand for the drug could be met with just 200 square kilometres of tobacco fields — an area smaller than the city of Boston. *eLife* 5, e13664 (2016)

DEVELOPMENTAL BIOLOGY

The likely root of night vision

The cells in the retina that enable night vision may have evolved from those that sense colour.

Typically, most of the light-sensing cells in mammalian retinas are rod cells, which are sensitive in low light. However, vertebrate ancestors only had cells resembling cones, which function under bright light and can discriminate colour.



EVOLUTIONARY BIOLOGY

Scales and fur have shared origin

Mammals, birds and reptiles inherited key cell structures that give rise to their fur, feathers and scales from a shared reptilian ancestor.

Scientists have long debated whether these skin appendages evolved independently or had a single origin. To find out, Nicolas Di-Poi and Michel Milinkovitch at the University of Geneva in Switzerland studied skin development in embryos of Nile crocodiles (*Crocodylus niloticus*; pictured right), corn snakes (*Pantherophis guttatus*) and bearded dragons (*Pogona vitticeps*). They found that reptilian scales, like

feathers and mammalian hair (mouse embryo pictured left), develop from a group of cells called the anatomical placode (pictured as dark blue spots). These appear only briefly during development in snakes and lizards, and were previously not detected and so thought to be missing in these animals.

These cells express the same developmental genes as bird and mammalian placodes, suggesting a common origin for modern hair, feathers and scales.

Sci. Adv. 2, e1600708 (2016)

ENERGY

Nanopores harvest wasted heat

A membrane with nanometre-sized pores can capture low levels of heat energy to generate power.

Industrial plants are abundant sources of waste heat, but the relatively small temperature difference between the source (which is usually below 100 °C) and its surroundings makes it hard to exploit. Menachem Elimelech of Yale University in New Haven, Connecticut, and his colleagues used a

water-repellant membrane that traps air in its pores and placed it between hot and cold water streams, creating a tiny air gap between the streams. The hot water evaporates on one side of the membrane, passes through the pores and condenses in the cold stream, creating hydraulic pressure that drives a turbine.

With a heat source at a temperature of only 60 °C, the device transferred power densities of up to 3.5 watts per square metre to a 20 °C fluid. *Nature Energy* <http://dx.doi.org/10.1038/nenergy.2016.90> (2016)

Ted Allison at the University of Alberta in Edmonton, Canada, Anand Swaroop at the US National Eye Institute in Bethesda, Maryland, and their co-workers studied mouse rod and cone cells, and monitored these cells in the developing mouse retina. They found that early in rod cells' development the cells expressed key genes that are normally active in 'S' (blue) cones. Zebrafish rods, however, did not.

The adaptation of cone cells to function under low light may have allowed mammals to adopt nocturnal lifestyles during mammalian evolution. *Dev. Cell* 37, 520–532 (2016)

A. C. TZIKA, N. DI-POI & M. C. MILINKOVITCH

IMMUNOLOGY

Insect bites make viral disease worse

Bites from mosquitoes that spread viruses trigger a distinct immune response in the skin after they bite, which increases the severity of infection caused by the transmitted virus.

Clive McKimmie at the University of Leeds, UK, and his colleagues injected mice with one of two mosquito-borne viruses. Mice that were bitten by virus-free mosquitoes and then injected with the microbe showed an immune response that retained more virus at infection sites than did infected mice that had not been bitten. Immune cells called neutrophils were drawn to the bite, where they enhanced the virus's ability to infect and multiply, causing more-severe disease.

Blocking certain immune cells from reaching the site of an insect bite reduces viral replication and could be a way to diminish disease after a bite, the authors say.

Immunity 44, 1455–1469 (2016)

BEHAVIOUR

Older monkeys socialize less

Like humans, some monkeys show declining social activity with age.

Laura Almeling at the German Primate Center in Göttingen, Germany, and her colleagues studied Barbary macaques (*Macaca sylvanus*; pictured), and found that older females spent less time grooming others and interacted with fewer animals than

younger individuals did. These changes were not explained by an overall reduction in social interest, as older males and females maintained an interest in pictures of other animals. Moreover, the monkeys' interest in toys and other non-social objects decreased in early adulthood, mirroring humans' declining eagerness for new experiences with age.

People's tendency to shrink their social circles as they age has previously been attributed to a sense that time is growing short, but the results in monkeys suggest that it may also be rooted in primate evolution, the authors say. *Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2016.04.066> (2016)

PLANT BIOLOGY

African trees cope with warming

Some trees in Africa already seem to be adapting to the warming climate by using water more efficiently.

Iain Robertson of Swansea University, UK, and his colleagues collected a small number of samples from three tree species in Ethiopia, Namibia and South Africa, covering a small area of the continent. By measuring the ratio of carbon isotopes in each tree ring, the team estimated the water-use efficiency of the trees from 1909 to 2003. They found that two of the three species increased their water-use efficiency — by an average of 25% — over the period.

Using water more sparingly may help to compensate for the predicted decreases in rainfall in Africa, allowing some plants to cope better with climate change than others.

J. Quaternary Sci. 31, 386–390 (2016)

CANCER BIOLOGY

T cells target solid tumours

A cancer therapy that uses genetically modified versions of patients' immune cells to

treat blood cancers has been adapted to attack solid human tumours implanted into mice.

Engineered T cells are designed to home in on specific proteins on the surface of cancer cells in the blood — but adenocarcinomas, a common type of solid tumour, rarely carry such markers. Avery Posey and Carl June of the University of Pennsylvania in Philadelphia and their colleagues developed a way to modify human T cells so that they recognize abnormal forms of a sugar molecule linked to a cell-surface protein that is abundant in many cancers. The authors found that in a mouse model of human pancreatic adenocarcinoma, all animals treated with these T cells survived until the end of the experiment, compared with only 40% of untreated controls.

Protein-linked sugars are a promising target for cancer immunotherapy, the team says. *Immunity* 44, 1444–1454 (2016)

EVOLUTION

When pupfish got to Devils Hole

A rare fish species living in an isolated cavern pool probably originated when the cavern first opened to the surface around 60,000 years ago.

The Devils Hole pupfish (*Cyprinodon diabolis*) is one of the world's rarest animals, and researchers debate whether humans introduced the fish to the pool in Devils Hole in the southwestern United States between 20,000 and 10,000 years ago. İsmail Sağlam and Michael Miller at the University of California, Davis, and their colleagues analysed the genomes of the fish and two related pupfish species, and concluded that the Devils Hole pupfish became an isolated population in the cavern roughly 60,000 years ago.

A geological event may have both opened up the cavern and introduced the pupfish into it, the authors suggest.

Mol. Ecol. <http://doi.org/bj78> (2016)



MATERIALS

Self-folding mimosa mimic

A bilayered material can curl itself into a cylinder in response to a stimulus, mimicking the leaves of the plant *Mimosa pudica*, which quickly fold up when lightly touched (pictured).

Zuankai Wang at the City University of Hong Kong, Antonio Tricoli at Canberra's Australian National University and their team were inspired by the plant. They adhered a hydrophobic layer, polyvinyl chloride, to a hydrophilic one, polycaprolactone, then placed this bilayer on a flexible plastic substrate before cutting the resulting trilayer into a long, thin strip. When they placed a water droplet on one end of the hydrophilic side, the two sides of the strip quickly peeled away from the substrate and wrapped around the droplet. As the water spread down the strip, the bilayer's edges curled with it to form a tube.

Such a material, which can be cut into different shapes, could one day be useful in sensors and other devices that don't require power.

Sci. Adv. 2, e1600417 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

RESEARCH

CRISPR human trial

The first human therapy to involve the CRISPR–Cas9 gene-editing technology passed a major hurdle on 21 June, when a federal advisory panel at the US National Institutes of Health approved a proposal to use the technique to edit T cells, a type of immune cell, taken from people with cancer. The trial, which would be run by the University of Pennsylvania in Philadelphia, would simultaneously enhance the T cells' ability to destroy cancerous cells and protect them from being attacked by the cells. US regulators have yet to approve the trial. See page 590 and go.nature.com/28qkj6m for more.

FACILITIES

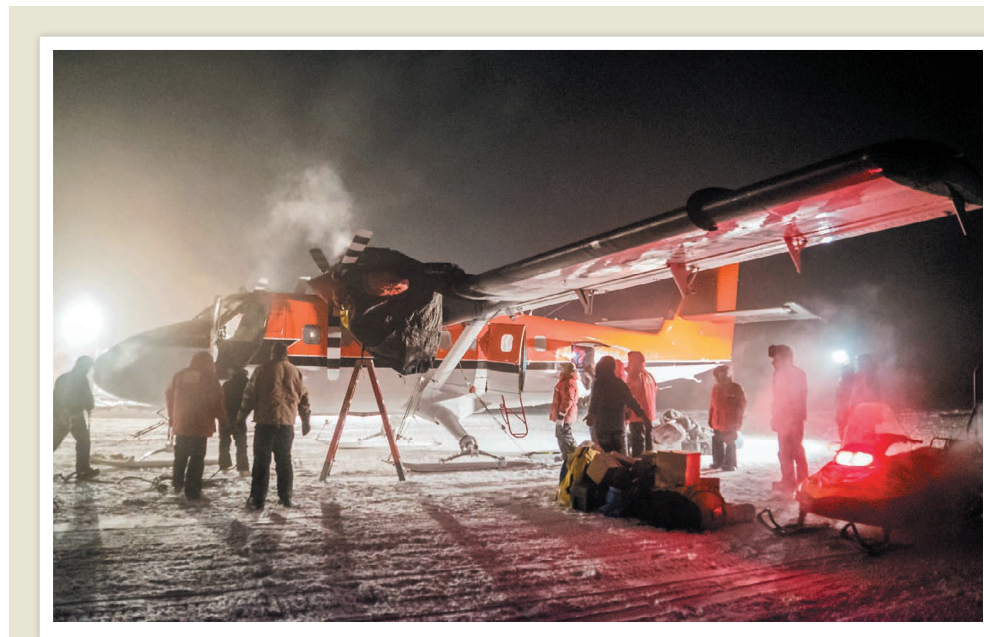
Olympic doping lab

The World Anti-Doping Agency (WADA) has suspended the accreditation of the laboratory in Rio de Janeiro, Brazil, that was to have handled anti-doping tests of urine and blood samples from athletes at the city's upcoming Olympic Games. WADA announced on 24 June that the facility had failed to conform with its international laboratory standards — but did not specify why. Brazil has had previous such troubles: it lacked a WADA-accredited lab for the Rio-hosted 2014 football World Cup. Football's governing body FIFA decided to fly samples to a lab in Switzerland for testing.

EVENTS

Brexit shock

The United Kingdom's vote to leave the European Union in a referendum on 23 June has left researchers scrambling to protect their scientific relationships and funding



ROBERT SCHWARZ/NATIONAL SCIENCE FOUNDATION

Two workers rescued in Antarctic mission

Two ill crew members were evacuated from the US Amundsen-Scott South Pole Station on 22 June. A Twin Otter aeroplane operated by Kenn Borek Air of Calgary, Canada, travelled to the pole after stopping at Britain's Rothera station. It was only the third midwinter flight ever made

to the pole, following medical evacuations in 2001 and 2003. The National Science Foundation, which oversees US research at the pole, did not release the names or conditions of the patients; both were flown to southern Chile and onwards to receive medical treatment.

streams. In a surprise to many observers, 52% of voters chose to leave the EU. In the run-up to the referendum, a number of senior academics and research organizations (and *Nature*) had voiced fears that a vote to leave would be highly disruptive to science. See pages 589 and 597 for more.

Coral crisis

More than 2,500 coral-reef scientists, policymakers and stakeholders have written to the Australian prime minister demanding that the government stop approving new coal mines, because climate change is the major threat to reef ecosystems. The letter, sent on 25 June after last week's International Coral

Reef Symposium in Honolulu, Hawaii, notes that Australia's Great Barrier Reef has been devastated by bleaching this year. Reef bleaching around the world will worsen as global temperatures rise. The signatories say that the government should "stop endorsing the export of coal" and halt plans for controversial mines in Queensland.

India space record

India's space agency set a record on 22 June by launching 20 satellites into orbit in a single mission — the biggest number in the agency's history. Its previous record for a single launch was ten satellites. The payload, which launched from a site in the eastern state of Andhra Pradesh, included

13 satellites from the United States. The achievement brings the agency's delivery rate closer to those of NASA and Roscosmos, and cements India's place as a major player in the space industry.

Chinese rocket

China's new Long March 7 rocket made a successful maiden flight on 26 June. The rocket, which launched from Hainan Island, is eventually intended for use in transporting cargo and people to a new Chinese space station planned for 2022. It uses a kerosene and liquid-oxygen fuel, which is less toxic than propellants of older Chinese rockets. The launch delivered several satellites to low-Earth orbit.

FERMILAB

PEOPLE

Trachea surgeon

Controversial surgeon Paolo Macchiarini, who pioneered transplants of artificial windpipes seeded with patients' own stem cells, is facing preliminary charges of involuntary manslaughter in connection with two patients who died after surgery, public prosecutors in Stockholm announced on 22 June.

Macchiarini is also suspected of causing grievous bodily harm to another transplant patient and to a patient undergoing a different type of operation, they said. In March, Macchiarini was fired from the Karolinska Institute in Stockholm — where he had worked since 2010 — after allegations of clinical and scientific misconduct. No formal charges have been brought and Macchiarini denies any wrongdoing.

Helen Edwards dies

Physicist Helen Edwards, a driving force behind the Tevatron particle accelerator at Fermilab near Chicago, Illinois, died on 21 June, aged 80. Edwards (pictured) led the design and construction of the Tevatron, which began smashing together protons and antiprotons in 1985; a decade later, observations of these collisions resulted in the discovery of the top



quark. Edwards also worked on accelerator designs for future high-energy-physics machines. The Tevatron closed in 2011.

AWARDS

Blavatnik awards

The three winners of this year's US Blavatnik Awards for Young Scientists were announced on 21 June. David Charbonneau at Harvard University in Cambridge, Massachusetts, was honoured for his work on observational astronomical methods used to search for chemical signatures of life in space. Phil Baran at the Scripps Research Institute in La Jolla, California, won for his research on the use of chemical synthesis to design scalable, efficient routes to potential new drugs. Michael Rape at the University of California, Berkeley, was rewarded for his discoveries

in cellular signalling involving the protein ubiquitin. Each person receives US\$250,000 — the largest unrestricted cash prize for early-career scientists. The prizes are awarded annually by the Blavatnik Family Foundation and the New York Academy of Sciences.

POLICY

Looser drone rules

The United States has markedly relaxed its rules that govern the use of small drones, clearing the way for commercial — and many scientific — applications. The policy, announced by the White House on 21 June, had been under development at the Federal Aviation Administration (FAA) for years. Many scientists had been unable to use drones for research because the machines could not be flown for 'commercial' use, which included research and teaching activities at private universities. The latest rules, which apply to drones weighing less than 25 kilograms, require commercial operators to be certified with the FAA. Drones must be kept within the line of sight.

Chemical control

Long-awaited reforms to US chemical regulations were signed into law on 22 June by President Barack Obama.

COMING UP

7 JULY

A Soyuz rocket launches to take Anatoly Ivanishin, Kate Rubins and Takuya Onishi to the International Space Station.

27 JUNE–2 JULY

The Starmus festival in the Canary Islands, Spain, brings together astronomy, art and music with speakers including Brian May, Stephen Hawking and Brian Eno.

www.starmus.com

The update to the 1976 Toxic Substances Control Act gives the US Environmental Protection Agency greater authority to ensure the safety of chemicals — both old and new. Under the revised law, the agency can request more information from chemical manufacturers and even compel firms to conduct extra safety studies. Several previous attempts to overhaul the law had failed over the past decade.

NASA travel ban

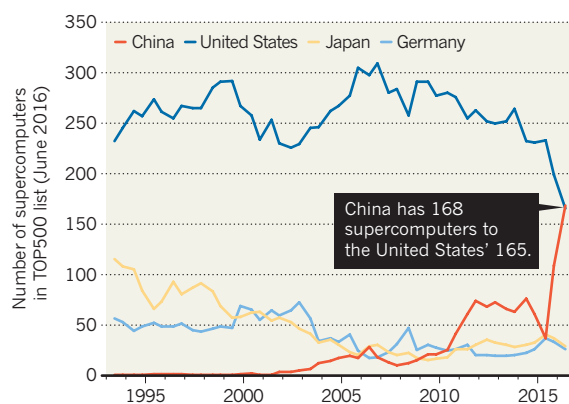
NASA has effectively banned its employees and contractors from attending a major space-research conference that begins in Istanbul, Turkey, on 30 July, citing security concerns. An internal memo dated 21 June reports that NASA head Charles Bolden made the decision to not sponsor or process travel to the Committee on Space Research (COSPAR) assembly in line with travel warnings issued by the US state department. Lennard Fisk, a space scientist at the University of Michigan in Ann Arbor and president of COSPAR, decried the decision as giving in to terrorist threats.

TREND WATCH

A Chinese computer tops the list of the world's 500 fastest supercomputers, for the seventh consecutive time. The leading machine, Sunway TaihuLight at the National Supercomputing Centre in Wuxi, can make 93 quadrillion calculations per second. It is almost three times as powerful as the previous list's winner, Tianhe-2, also in China. For the first time, China overtakes the United States in number of supercomputers in the biannual TOP500 ranking. It had just one machine on the list until 2000.

SUPERCOMPUTER SUPERPOWER

China has ended the United States' dominance in supercomputing, overtaking it in number of machines, as well as speed.



SOURCE: TOP500

NATURE.COM

For daily news updates see:

www.nature.com/news

NEWS IN FOCUS

ASTRONOMY NASA's Juno spacecraft set to buzz past Jupiter **p.599**

CLIMATE CHANGE Glacial growth shrinks Indus River flow **p.600**

PUBLISHING ArXiv preprint server mulls makeover **p.602**



CLONING The day we made Dolly the sheep **p.604**

NEIL HALL/REUTERS



The UK vote to leave the European Union has sparked huge uncertainty across the continent.

POLITICS

UK scientists in limbo after Brexit shock

Researchers organize to lobby for science as country prepares for life outside the EU.

BY ALISON ABBOTT, DANIEL CRESSEY AND RICHARD VAN NOORDEN

The dust from last week's vote by the United Kingdom to leave the European Union is nowhere near settled, but the country's researchers are already bracing for the fallout.

On 23 June, 52% of those who voted in the country's referendum came out in favour of

leaving the EU. No one is sure how 'Brexit' will affect science, but many researchers are worried about long-lasting damage. Beyond the immediate economic impacts and the potential loss of EU funding — which currently supplies some 16% of UK university research money — scientists fear a loss of mobility between the country and the continent.

"I was on a career panel only yesterday, singing the praises of the UK as a wonderful place

of opportunity for young scientists, and I feel like that has changed overnight," said Vanessa Sancho-Shimizu, an infectious-diseases researcher at Imperial College London, in response to a *Nature* survey last Friday. She is a Spanish national and one of many scientists who expressed similar views.

Researchers are already mobilizing to lobby for the United Kingdom to remain a participant in EU science programmes, and for ►

► domestic funding to make up any shortfalls. “We need some kind of rapid monitoring to catch fallout problems early and implement remedial measures,” says Mike Galsworthy, who led the Scientists for EU campaign.

“If the science community wants to have an impact on the UK’s negotiation strategy, it needs to clearly know what its own priorities are and start the process of making that case, strongly,” says John Womersley, chief executive of the UK Science and Technology Facilities Council. Getting a guarantee to remain part of Horizon 2020, the EU’s €74.8-billion (US\$82.9-billion) programme of research grants, should be the community’s top — and only — objective, he adds.

Jamie Martin, an independent education consultant who advocated for Brexit, offers “total reassurance” to worried scientists. Most academic groups had lobbied for the United Kingdom to remain in the EU. Martin says that “the good news for them is that the people at the top of the Vote Leave campaign share their instincts on science”. This includes being open to skilled people from other countries and understanding the importance of continued funding, he says.

PEOPLE

Exactly when the United Kingdom will leave the EU is unclear. There is no set date for the government to invoke ‘article 50’ of the EU Lisbon treaty, but once it does, it will trigger a process of negotiation that must conclude within two years. Campaigners for a Leave vote — including former London mayor Boris Johnson, whom many expect will lead the next government — have said that there is no need to do this immediately, and informal negotiations with the rest of the EU can take place first.

Those in favour of Brexit say that a United Kingdom outside the EU could allow in more skilled researchers while still driving down overall immigration numbers. ‘Leave’ campaigners have advocated a points-based immigration system such as Australia’s, which would attempt to level the playing field between EU and non-EU researchers.

But it’s unclear whether the United Kingdom will still be attractive to talented researchers. Some have said that they feel less welcome in the country as a result of both the vote and the campaign leading up to it, which featured highly charged rhetoric around immigration.

MONEY

Even laboratories staffed primarily by UK nationals could feel the pinch. EU research funds have supplied an estimated €8 billion



Mike Galsworthy wants careful monitoring of UK research to spot any fallout.

to the country over the past decade.

The United Kingdom is also by far the largest recipient of loans to EU universities and research institutions from the European Investment Bank (EIB), receiving more than €2.8 billion since 2005 — some 28% of total EIB loans for higher education and research over that period. Agreed loans are secure, but the fate of those that are just beginning to be considered is unclear, says EIB spokesman Richard Willis.

Leading campaigners for the Leave side pledged before the vote that universities and scientists in the United Kingdom who now get funding from the EU “will continue to do so”.

“The long-term future worries the hell out of me.”

EU countries currently hold within Horizon 2020. But that might not be possible if the country acts to restrict free movement of people, as many Leave supporters have demanded. Switzerland, a non-EU member, is an associated country, but its researchers were cut out of full access to Horizon 2020 after the nation voted in a 2014 referendum to restrict immigration.

“The long-term future worries the hell out of me,” says Steven Cowley, who directs the Culham Centre for Fusion Energy in Abingdon, UK. The centre operates the Joint European Torus (JET), a nuclear-fusion facility, on behalf of the European Commission. The contract for JET runs out in 2018, but Cowley says he is confident that it will be extended, because it provides crucial expertise for ITER, the international fusion experiment

The country could try to negotiate access similar to the agreements that 15 other non-

under construction in southern France. The real problem, he says, is that the United Kingdom will not be able to compete to host the next major European facility.

As for ITER itself, the EU is one of seven major international members of the project. The United Kingdom will have to rejoin it, either as an individual nation member — which would mirror its membership of CERN, the European particle-physics lab — or perhaps with ‘associate member’ status similar to that held by Switzerland.

POLICY

A UK exit from the EU could also reshape the policy landscape for the countries that remain in the bloc.

Germany, Italy and Austria are among the nations that have opposed EU funding for research on human embryonic stem cells.

Others, including the United Kingdom and Sweden, called for research to be funded under appropriate ethical oversight — leading to a deal in which research collaborations can be funded as long as partners from countries where the research is forbidden do not handle human embryonic stem cells themselves. The United Kingdom was “in the forefront of guiding us into an acceptable and workable way around the issues”, says stem-cell researcher Christine Mummery of the Leiden University Medical Center in the Netherlands. “If the UK cannot participate in decisions like this, it makes me nervous.”

Other European scientists fear for the future of their own countries’ science bases if the UK vote empowers other anti-EU movements. Right-wing populist politicians in France, the Netherlands and Denmark are already calling for their own referendums.

James Wilsdon, a science-policy researcher at the University of Sheffield, UK, says that beyond the questions about continued access to EU funding and policy, there is a more fundamental issue that UK researchers must come to grips with: the fact that most academic experts, research lobby groups and other experts came out in favour of staying in the EU and were ignored by the public.

“Here you have such a major question around which there was such a torrent of solid analysis and empirical evidence, and we’ve had a rejection of that by 52% of the public,” he says. “That needs to provoke some serious soul searching and reflection.” ■

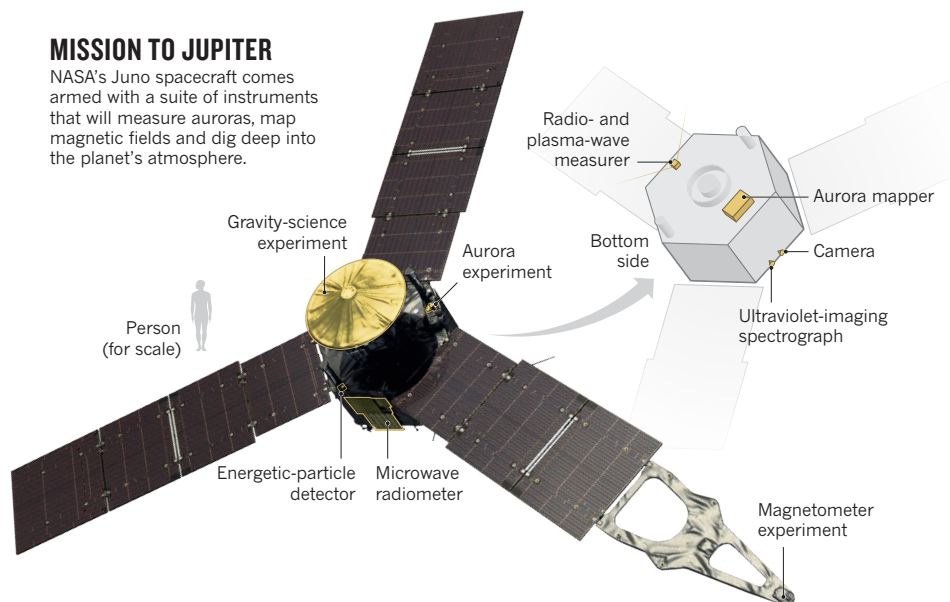
SEE EDITORIAL P.589

Additional reporting by Davide Castelvecchi and Elizabeth Gibney.

GAVIN BLACK PHOTOGRAPHY

MISSION TO JUPITER

NASA's Juno spacecraft comes armed with a suite of instruments that will measure auroras, map magnetic fields and dig deep into the planet's atmosphere.



PLANETARY SCIENCE

NASA spacecraft nears Jupiter

Juno will explore the gas giant's composition and mysteries.

BY ALEXANDRA WITZE

On 4 July, NASA intends to finish a job that started with the agency's Galileo mission 21 years ago. At 8:18 p.m. Pacific time, the Juno spacecraft will ignite its main engine for 35 minutes and nudge itself into orbit around Jupiter. If all goes well, it will eventually slip into an even tighter path that whizzes as close as 4,200 kilometres above the planet's roiling cloud-tops — while dodging as much of the lethal radiation in the planet's belts as possible.

The US\$1.1-billion mission, which launched in 2011, will be the first to visit the Solar System's biggest planet since NASA's Galileo spacecraft in 1995. Picking up where Galileo left off, Juno is designed to answer basic questions about Jupiter, including what its water content is, whether it has a core and what is happening at its rarely seen poles (see 'Mission to Jupiter').

Scientists think that Jupiter was the first planet to condense out of the gases that swirled around the newborn Sun 4.6 billion years ago. As such, it is made up of some of the most primordial material in the Solar System. Scientists know that it consists mostly of hydrogen and helium, but they are eager to pin down the exact amounts of other elements found on the planet.

"What we really want is the recipe," says Scott

Bolton, the mission's principal investigator and a planetary scientist at the Southwest Research Institute in San Antonio, Texas.

A MURKY DISPOSITION

Jupiter's familiar visage, with its broad brown belts and striking Great Red Spot, represents only the tops of its churning clouds of ammonia and hydrogen sulfide. Juno — named after the Roman goddess who could see through clouds — will peer hundreds of kilometres

into the planet's atmosphere using microwave wavelengths.

Exploration of Jupiter's interior should reveal more about the formidable atmospheric convection that powers the planet, says Paul Steffes, an electrical engineer at the Georgia Institute of Technology in Atlanta.

Steffes and his colleagues have run a series of laboratory experiments to simulate what different layers of Jupiter's atmosphere might look like: from near the cloud-tops, where experimental temperatures are -100°C to deeper in the planet, where they rise to more than 300°C .

By comparing Juno's observations to their simulations, the scientists hope to determine how much ammonia, water vapour and other materials swirl at different atmospheric depths. "Once we understand the recipe for Jupiter's atmosphere, we'll get a clearer insight into how it evolved," says Steffes. Different theories predict varying amounts of water in Jupiter's atmosphere, depending on whether the planet coalesced at its current distance from the Sun or somewhere else. Actual measurements of atmospheric water content could help to clarify this debate.

NORMAL IS GOOD

In anticipation of Juno's arrival, professional and amateur astronomers have been observing Jupiter with ground-based and space-based telescopes. For now, the planet is not experiencing any unusual atmospheric changes. "It's kind of in its normal state, which is good," says Amy Simon, a planetary scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. This 'normal' behaviour gives researchers confidence that they will be able to understand Juno's findings.

The Great Red Spot continues to shrink, as it has done in recent years, and to interact less and less with the jet streams on either of its edges. The broad belt just north of the planet's equator has been expanding since late 2015 ▶



Jupiter's Great Red Spot also reveals a mosaic of currents that swirl through the planet's atmosphere.

► — a change that might be connected to processes deep in the atmosphere.

“Trying to connect events that are happening at one level to events happening in another tells you how well coupled the whole atmosphere is,” says Leigh Fletcher, a planetary astronomer at the University of Leicester, UK.

As Juno probes deeper and deeper into the planet’s atmosphere, researchers hope to get information on a layer of hydrogen compressed into a liquid by increasing pressures. That liquid conducts electricity, which powers Jupiter’s enormous magnetic field. Deeper still, the spacecraft will look for evidence of a core — a dense nugget of heavier elements that most scientists think exists, but has never been observed. Juno will make precise measurements of how Jupiter’s gravity tugs on the spacecraft, which should reveal whether a core is present.

POLE POSITION

Juno will also get an unprecedented glimpse of Jupiter’s poles. To avoid the most dangerous radiation belts that surround the gas giant — which over the lifetime of the mission could fry the spacecraft with the equivalent of more than 100 million dental X-rays — Juno will take a long elliptical dive around the planet on every orbit. The spacecraft will fly directly over Jupiter’s magnetically intense auroras, and could spot unusual circulation patterns that resemble a hexagon-shaped feature parked on Saturn’s north pole.

The lessons that scientists learn from Jupiter will apply to other gas giants, including those outside the Solar System. “If we understand how it formed, we’ll have a much better handle on giant-planet influences in planetary systems around other stars,” Fletcher says.

Juno will provide scientists’ last chance to look at Jupiter for a long time. It is scheduled to make 37 total orbits before performing a kamikaze run in early 2018, burning up inside the planet’s clouds to keep it from contaminating the moon Europa. The only other mission planned to the gas giant is the European Space Agency’s Jupiter Icy Moons Explorer (JUICE) spacecraft, which could launch as early as 2022 and will focus mainly on the moon Ganymede. ■



Population growth and agriculture have stressed the Indus, which flows the length of Pakistan.

CLIMATE CHANGE

Indus River waters shrinking

Cooler, cloudier summers slow snowmelt in Himalayas.

BY JANE QIU

The Indus River, which supports the lives of 300 million people, is supplying Pakistan with less water than it did 50 years ago, particularly in the spring and summer, researchers have found. The news comes as demand for water is projected to rise sharply.

The findings contradict previous predictions that the river’s volume would stay the same, or even grow, as climate change kicks in, although that increase is likely to occur in the next several decades, another team has found.

Danial Hashmi, a hydrologist at the Pakistan Water and Power Development Authority in Lahore, reported the river’s shrinkage for the first time in February at a conference in Kathmandu. Further data from India have also shown seasonal shifts. “The Indus is certainly

changing, and local communities are feeling the pinch,” Shresth Tayal, a glaciologist at the Energy and Resources Institute in New Delhi, told a meeting in Columbus, Ohio, last month.

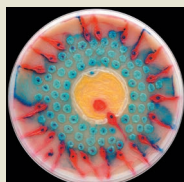
The Indus flows through India, Afghanistan and China before reaching Pakistan, which it crosses from north to south. For decades, population growth and agriculture have stressed the river, which, for 10 months of the year, dries up before it reaches the sea. Because demand is set to rise by 30% by 2025, “water shortage will be the single most destabilizing factor, not only for Pakistan but the entire region,” says Arif Anwar, principal researcher at the International Water Management Institute in Lahore.

But since the 2009 ‘glaciagate’ scandal — in which it emerged that the Intergovernmental Panel on Climate Change had mistakenly included in its fourth assessment report



**MORE
ONLINE**

IMAGES OF THE MONTH



Nature’s picks of May’s top science photos
go.nature.com/2985ozy

MORE NEWS

- Scientists atwitter over United Kingdom’s Brexit vote go.nature.com/295fgqh
- Rhino relocation to Australia under fire go.nature.com/2911gxx
- First ‘nonlinear’ simulations of cosmic expansion go.nature.com/291x6ti

NATURE PODCAST



Dolly the sheep’s legacy; the trials of funding interdisciplinarity; and a social science ‘IPCC’ nature.com/nature/podcast

a prediction that the Himalayan glaciers would disappear by 2035 — there has been a widespread belief that water resources in the region are stable, at least for now. Research by several groups even suggested that climate change might provide some relief in the short or medium term, thanks to faster melting of the glaciers that supply the river, and increased precipitation.

Hashmi's data, which are unpublished, come from a network of hydrological stations in Pakistan that span the main stem of the Indus and three of its tributaries. They show that the total water supply fell by 5% between 1962, when the hydrological stations were built, and 2014.

"A reduction of 5% over five decades may not seem a lot," says Walter Immerzeel, a hydrologist at Utrecht University in the Netherlands, who led one of the studies that projected an increase in water supply in the Indus (A. F. Lutz *et al. Nature Clim. Change* 4, 587–592; 2014). "But if the trend persists, there could be devastating implications for water resources."

Hashmi's team finds that the river's shrinkage is seasonal, with a decrease in flows between April and August that exceeds a slight increase during the rest of the year. And it reports a temperature drop across the four Pakistani river basins in the summer months — even though the region is getting warmer overall. Because snow- and glacier

melt contribute to 50–85% of river flow in those catchments, the team suspects that cooler springs and summers result in less melt and that this can explain the shrinking river.

"It's a fascinating finding," says Tobias Bolch, a glaciologist at the University of Zurich in Switzerland. He notes that it is consistent with a phenomenon known as the Karakoram anomaly, in which some of the glaciers in the region

"If the trend persists, there could be devastating implications for water resources."

have become stable or even grown — in contrast to most mountain glaciers globally, which are retreating rapidly in response to climate change.

Another study presented at the February meeting suggested a possible reason for the region's cooler summers. As the overall climate warms, monsoons increasingly invade the mountain chains of the Indus upstream, where glaciers reside, says study co-author Hayley Fowler, a climate modeller at Newcastle University, UK. Her modelling work shows that when monsoons penetrate into the region and push dry westerly winds northward, summer temperatures drop. The team suspects that monsoonal clouds hovering over a region that is normally hot and dry in the summer may have a cooling effect.

The limitations of climate models and the scarcity of field measurements in the region make it hard to predict how Himalayan water resources will change, says Immerzeel. However, the latest work by him and his collaborators — which took the Pakistani data into account — finds that things will get much worse, but only in the long term. Using state-of-the-art climate models, and assuming a scenario in which global greenhouse-gas emissions peak around 2040, the team found that the flow of water in the river system will stabilize or even increase in the next few decades — consistent with its previous results. But once glaciers have become depleted and regional temperatures have started to rise, water scarcity will ensue: the researchers predict a 15% drop between 2071 and 2100 compared with 1971–2000 levels, Immerzeel says. The team has submitted a paper for review.

In any case, there is a pressing need for Pakistan to boost its water-storage capacity and efficiency of water usage, says Mobin-ud-Din Ahmad, a hydrologist at the Commonwealth Scientific and Industrial Research Organisation in Canberra, Australia. Right now, its reservoirs can hold only 30 days' worth of the country's water needs — compared with 800 days in Australia and 150 days in India. "It's an extremely dangerous situation, especially now, when severe droughts are increasingly common," he says. ■

Preprint website plans revamp

But users are wary of major changes to arXiv repository.

BY RICHARD VAN NOORDEN

A multimillion-dollar funding drive is being readied to transform arXiv, the vastly popular repository to which physicists, computer scientists and mathematicians flock to share their research preprints openly.

But the results of an enormous user survey published this week suggest that researchers are wary of drastic changes to a site that has become an essential part of the infrastructure of modern science.

Last year, the site served up around 139 million downloads, and it now holds more than 1.1 million free papers. But it is being sustained by fragile code, donations from libraries and a charitable foundation and the good will of about 150 or so volunteer moderators, says the site's programme director, Oya Rieger. With its 25th anniversary approaching in August, arXiv's advisory teams of scientists and librarians are considering a plan that involves raising US\$2.5 million to \$3 million to modernize the platform. That will sit on top of its \$1-million annual budget for staff and servers.

To attract support from donors, arXiv's operator, Cornell University Library in Ithaca, New York, is hoping to come up with a "compelling vision", Rieger says.

Scientists seem to love arXiv: 95% of the survey's 36,000 respondents said that they were very satisfied or satisfied with it. And most want to keep it just the way it is, although perhaps with some modernization. They were enthusiastic about the possibility of tweaks to

improve the site's search functions, and about allowing references to be hyperlinked directly to research papers, for example (see 'What do arXiv users want?'). Some wanted the site to broaden into new subject areas, such as chemistry — although such expansion would require the recruitment of scientists who are willing to moderate the manuscripts, notes David Morrison, chair of arXiv's scientific advisory board.

SOCIAL FORUM

When asked whether arXiv should embark on more transformational changes, respondents gave mixed answers. In particular, some questions focused on whether it should develop into a social forum that allows scientists to comment on papers or leave ratings. A few social-media sites have already been built around the repository for just such purposes — such as SciRate and Arxiv Sanity Preserver — and some argue that the site itself should begin to incorporate such functionalities. "ArXiv should be more dynamic — allowing readers to filter the wheat from the chaff," says Alán Aspuru-Guzik, a quantum chemist at Harvard University in Cambridge, Massachusetts. But one-third of respondents said that this wasn't important or that arXiv shouldn't be doing it. Only 34% voted in favour of such changes.

That response points to a tension between

researchers who want to see the site incorporate aspects of open review, and those who want it to stick to its core mission of allowing rapid exchange of scholarly papers, says Rieger. There were hints of a generational divide, with those aged under 30 more in favour of allowing comments. But even those who wanted a more social site said that they were keen to avoid a commenting free-for-all, Rieger adds.

"The message was more or less 'stay focused on the basic dissemination task, and don't get distracted by getting overextended or going commercial,'" says Paul Ginsparg, a physicist at Cornell University who launched arXiv in 1991 as a pre-World-Wide-Web-era bulletin board.

CHECKS AND BALANCES

Ginsparg notes, however, that arXiv's users sometimes don't know what they want until they get it. Researchers said that they liked the quality control now built into the site, including checks of papers for text overlap with other reports (potential plagiarism), classifying papers into the correct subject areas and rejecting work that has little scientific value. "These are for the most part things that users never actually requested," Ginsparg says. In the past 5 or so years, he has introduced automated machine-learning code that filters through the more than 9,000 papers submitted each month and flags up potential issues to human moderators.

In September, arXiv's advisory boards will meet to draw up a road map for progress and to discuss how to get the funds needed to modernize the site. The site is currently sustained by member institutions (mainly libraries, but also some research funding agencies) and by the Simons Foundation in New York. But some discussions have been held with other potential contributors such as the US National Science Foundation. It is also possible that publishers or scientific societies could be asked to contribute, says Rieger.

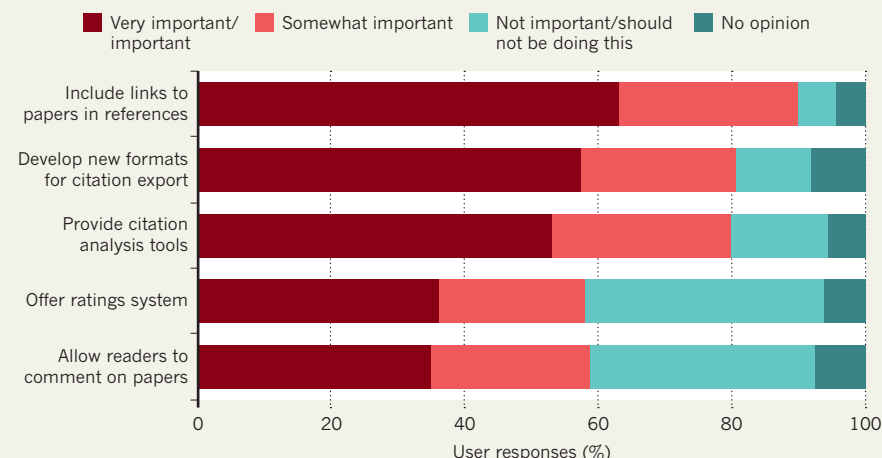
She adds that the site will need to be careful to remain objective. "We want to make sure that arXiv continues to be a neutral, trusted service," she says. ■

CORRECTION

The science-workforce graph in 'China by the Numbers' (*Nature* **534**, 452–453; 2016) erred in stating that China's population is 1.3 trillion. It is more than 1.3 billion.

WHAT DO ARXIV USERS WANT?

The preprint repository got high marks overall from 36,000 respondents to a survey, but there was no consensus over whether the site should add social-media functionalities.



Dolly was born
20 years ago.

On the day we made *Dolly*

BY EWEN CALLAWAY

The story of the world's most famous sheep, from the people who brought her to life.

THE CAST: **Karen Walker**, embryologist, PPL Therapeutics, Roslin, UK, now director, KXRegulatory, Linlithgow, UK; **Bill Ritchie**, embryologist, Roslin Institute, now at Roslin Embryology; **Angela Scott**, cell-culture technician, PPL, now chief operating officer, TC BioPharm, Motherwell, UK; **Alan Colman**, research director, PPL, now at Harvard University, Cambridge, Massachusetts; **Ian Wilmut**, embryologist, Roslin, now University of Edinburgh, UK; **John Bracken**, farm research assistant, Roslin, now retired; **Angelika Schnieke**, molecular biologist, PPL, now Technical University of Munich, Germany; **Harry Griffin**, scientific director, Roslin, now retired; **Jim McWhir**, stem-cell scientist, Roslin, now retired.

Dolly, the first mammal cloned from an adult cell, was born 5 July 1996. But she was created five months earlier, in a small room at the Roslin Institute, outside Edinburgh, UK.

Karen Walker: On the day we made Dolly, we had such a rubbish day.

Bill Ritchie: It was 8 February 1996. I looked it up. We do know it was a rubbish day: we had various problems with infections and things.

Walker: It's a shame the building has been demolished, otherwise you could see the room in which Dolly was made. I use the word 'room' loosely, because it really was just a big cupboard, which, when Bill and I were in there, you could just get two chairs and the incubator in.

Ritchie: It literally was the cupboard. It was the storage cupboard at the end of the lab. When we got camera crews in later, they couldn't believe it, there was no room to shoot.

Walker and Ritchie were part of a project at the Roslin Institute and spin-off PPL Therapeutics, aiming to make precise genetic changes to farm animals. The scientific team, led by Roslin embryologist Ian Wilmut, reasoned that the best way to make these changes would be to tweak the genome of a cell in culture and then transfer the nucleus to a new cell.

Ritchie: The simple way of describing nuclear transfer is that you take an oocyte, an unfertilized egg, and you remove the chromosomes. You then take a complete cell which contains both male and female chromosomes — all of our cells do, apart from the gonads. You take that cell and fuse it to the enucleated egg, activate it — which starts it growing — and transfer it to a surrogate mother. Hopefully, with your fingers crossed, you will get a cloned offspring, a copy of the animal you've taken that cell from.

Walker: Tedious is absolutely the word. You're sitting, looking down a microscope and you've got both hands on the micromanipulators. It's kind of like the joysticks kids use nowadays on games. If your elbow slipped, you could wipe the whole dish out.

A year earlier, the team had produced twin sheep, named Megan and Morag, by cloning cultured embryonic cells in an effort spearheaded by Roslin developmental biologist Keith Campbell. But on this day in February 1996, problems with the fetal cell lines they had planned to use meant that they would need another nuclear donor.

Walker: My memory is of flapping like a chicken, thinking, 'What are we going to put in?' because the cells we were going to use aren't there. The last thing you want to do is waste those oocytes you've got. We wanted to try something, at least.

Angela Scott: I received word from Karen to say that the cells they were expecting had been contaminated. They asked me if I had any cells that they could use. The cells I had were ovine mammary epithelial cells: we were looking to increase expression of proteins in milk. These were adult cells.

Alan Colman: I had come from a background of nuclear transfer with John Gurdon [a developmental biologist at the University of Cambridge, UK]. He'd never been able to get an adult frog by using nuclear transfer from an adult cell donor. He'd been able to get tadpoles using adult cells, but he'd never been able to get an adult frog. I didn't think it would work with adult cells at all. But we had no other cell line to go with, so we all agreed that we'd use these mammary-gland cells and just see what happened, gain some experience. These were from a 6-year-old sheep — middle-aged for a sheep.

➔ **NATURE.COM**
To watch an interview with the Dolly team, visit go.nature.com/28ouboua

Ian Wilmut: This is something that is got wrong to this day. Dolly is described as the first mammal cloned from an adult cell. She's actually the first adult clone, period. She's often undersold.

Although cloned and transgenic cows would be more valuable for industry, the Roslin team worked with sheep for practical reasons.

Wilmut: Cattle are incredibly expensive and have a long generation interval. Sheep are much less expensive and much easier to work with. And we knew the reproductive biology. It was very likely that if we could make something work in sheep, it would work in cows. Sheep are small, cheap cows.

John Bracken: There would be 40–60 animals going through surgery [to retrieve oocytes or implant embryos in surrogates] each week during the breeding season. It's a lot of different sheep in the system, and that had to be very accurately monitored so the animals were at the right place at the right time.

Walker: Bill used to keep the embryos and oocytes — when he was bringing them back up from the farm — in his top shirt pocket. I didn't have a top shirt pocket, so I used to tuck them inside my bra. It was a way to keep them warm and fetch them back into the lab and get them into a proper controlled environment. I don't think inside my bra was terribly controlled, but neither was Bill's top shirt pocket.

Ritchie: On the day we made Dolly, I would have done the enucleation, and she would have done the fusion. That was our normal way of doing things.

Walker: I did the fusion on the day we made Dolly. Bill and I joke, that he's the mum and I'm the dad because, essentially, I was the mimic to what the sperm would do.

They transferred 277 nuclei from the mammary cell line — from a white-faced breed known as a Finn Dorset — into eggs from the hardy Scottish blackface breed. Just 29 of the resulting embryos were implanted into surrogate ewes. Expectations were low: it seemed almost impossible that an adult cell nucleus could be reprogrammed to give rise to a live animal. Most cloned embryos aborted, many even before a pregnancy could be determined with ultrasound.

Wilmut: The sheep breeding season begins in October and ends in February, March-ish. By Christmas, we had established pregnancies after transfer from fetal cells, so that was going well. If we hadn't done that, we probably wouldn't have gambled on working with what became Dolly, the mammary cells.

Angelika Schnieke: I remember meeting

Ian Wilmut in the canteen, and he was very sceptical. He said: "I would be surprised if it works, but PPL is paying for the experiments, so we're doing them."

Bracken: We scanned all the recipients that had embryos transferred, and we knew they were important sheep. Every day that the scientists knew we were scanning, they would be very keen to know if there were any pregnancies.

Walker: I didn't go down to watch all the scans. But with Dolly — because we knew that those were cells Bill and I had put in — I had gone down on that particular day with John.

Bracken: I was just really pleased that it was a pregnancy. I didn't realize the real importance of it because we weren't really told. We just knew it was an important pregnancy. It didn't carry the same weight. We weren't thinking, 'Wow! If this progresses to a live lamb, this is going to be a world beater, or it's going to turn scientific understanding on its head.'

Walker: I'd taken a blank video up with me, so that I could show my colleagues. That video is sitting up in my loft, and to my shame, I have never yet transferred it onto DVD. I should.

Schnieke: I remember the day when we had the first scan. We always asked. And then we saw the picture and the scans. Then you just have to hope that it lasts and goes all the way through.

Wilmut: My memory is they were looking around day 30 or 35, so there's another 120 days [until the birth], where you keep on sighing with relief and hoping.

Just a few of the team members got to witness her birth.

Bracken: It happened about 4:30 in the afternoon. As soon as she went into labour, we called the Dick Vet [the Royal School of Veterinary Studies in Edinburgh] to get one of their vets to come out. Even though [farm research assistant] Douglas McGavin and myself probably had 50 years of experience between us, it just would have been unheard of if we'd decided we'd assist the birth and something had gone wrong.

Ritchie: We knew Dolly was about to be born, and I think she was showing signs of getting near lambing, and lo and behold I went through and there were bits of Dolly being born. There was a vet there, so she made sure the animal was okay and pulled the lamb out.

Bracken: It was absolutely normal. No complications whatsoever. She was a very viable lamb. She got on her feet very quickly, probably

Ian Wilmut with Dolly on display at the National Museum of Scotland.

"With hindsight, without a doubt it was a great name."

within the first half hour, which is a really good indication that things are normal.

Ritchie: I think I was jumping up and down when I saw that white face.

Scott: Karen was away at a wedding at the time.

Walker: I had given her the fax number of the hotel. I wish I had kept that fax. It said: "She has a white face and furry legs."

Scott: I don't know what they must have thought at the hotel: "Wow, that's a really unusual baby."

Wilmut: I was in the allotment. I had a phone call to say we had a live lamb. I issued an instruction that nobody should be there who didn't have to be there. Lots were curious. I obeyed my own rule because I'd got nothing to contribute.

Bracken: I'm standing next to Douglas McGavin watching the vet assist this birth, and I made an off-the-cuff remark to Douglas. I said, "You know what we're going to have to call this lamb? We're going to have to call it Dolly", after Dolly Parton, because the cells are derived from mammary tissue.

Wilmut: Being somewhat puritanical, I might

have been a bit worried. With hindsight, without a doubt it was a great name.

Bracken: This is hearsay. I never got told this directly. But I heard they had contacted Dolly Parton and said: "We've got this cloned sheep that's named after you."

Wilmut: I don't know how the message came through, but we were told her agent had said: "There was no such thing as baaad publicity." I don't know if that's true.

Over the next few months, Wilmut's team confirmed that Dolly was a clone of the mammary cell line, and wrote up the results. Her birth was to be kept top secret, until the *Nature* paper describing the experiment could be published in February 1997 (I. Wilmut et al. *Nature* 385, 810–813; 1997).

Harry Griffin: Two or three months before the publication of the paper, I got to know about it. In terms of preparation, PPL were involved. They saw it as an opportunity to get publicity for themselves. We worked with their PR company, De Facto. We did quite a bit of preparation.

Wilmut: Ron James, who was the chief executive of PPL therapeutics, and I were cited as the primary spokesmen and given a bit of training by ex-BBC people, who first of all came up and fairly aggressively stuck microphones up our noses and asked aggressive questions, and subsequently did it very gently. We weren't approached in anywhere near the aggressive way they tried first, which was quite shocking. I'm sure it was worth having.

Griffin: We had everything organized. The calls would be directed to De Facto and they would try and organize some coherence in our response in terms of who got priority and who didn't. All this would culminate, we hoped, on the Thursday that the paper came out. What was that, 27 February? Clearly, it didn't.

Wilmut: Robin McKie at *The Observer* leaked it. He will deny the charge.

Robin McKie, science and technology editor, The Observer, London: I didn't see that stuff in *Nature*. I don't blame him for being angry, but I went to great pains to avoid the things that would get me to be accused of that. I had helped a couple of guys who were making a TV programme about genetics, and they said, "Oh, by the way, they've cloned a sheep in Edinburgh." I didn't believe them, but I phoned a few people in the field, and one of them in America confirmed it. But I was very, very worried. I was saying something quite sensational, with absolutely no paper proof of

REUTERS

A Sunday news report about Dolly brought intense media attention to Scotland.



“It was just a forest of flash bulbs and reporters.”

JEFF J. MITCHELL/REUTERS

anything that had gone on. I told my deputy editor everything I knew, and he made me write it. Then the shit hit the fan.

Griffin: Ian gave me a call and said he'd just been called up and told that *The Observer* was going to run the story on the Sunday prior to publication in *Nature*.

Ian and I went into the institute at about 9 a.m. on the Sunday, not knowing whether or not people could get through. The phone rang continuously. We had a bizarre circumstance where a phone started ringing in a cleaning cupboard. When I answered it, it was, I think, the *Daily Mirror*, who had somehow got this particular connection. About half past nine at night, we went home.

Jim McWhir: I remember coming in on the day after the embargo broke and there were several satellite vans in the carpark.

Wilmut: There were television trucks everywhere. I went and spoke on *Good Morning America*.

Griffin: CBS, NBC, ABC, BBC, all there wanting interviews with Ian, wanting to see the sheep. It was chaos. I don't think you can ever appreciate the intensity of the media in full flight unless you've experienced it yourself.

McWhir: It was just pandemonium. Going down to the large-animal unit, it was just a forest of flash bulbs and reporters. It was quite amazing. I just turned around and went back to work.

Griffin: My secretary would put the phone down, and it was ringing immediately.

One of the names I heard being mentioned was Harold Shapiro [then chair of the US National Bioethics Commission]. She said, “Ian Wilmut can't talk to you now, can you call back later?” Bill Clinton had asked him to report back within 90 days on the ethical implications of cloning. I overheard his name, and said, “No, we definitely want to talk to him.”

Colman: When you're embedded in a project, you have what you consider to be good scientific reasons for doing it. Everything we did was covered by an ethics committee. We had been through a lot of concerns about animal health. Our concern was more about that kind of reaction. We weren't doing it as a prelude to cloning humans.

Griffin: People in the media pressed this point repeatedly. We were accused of keeping Dolly's birth secret because we were contemplating cloning a human. We had our position clear on that: it was unethical and unsafe.

Wilmut: It goes with the job. You just have to explain this is not the case.

Schnieke: In Europe, it was immediately seen as a negative. “What have they done now and

what could they do next?” We had police at the institute who explained what you do if there's a bomb scare. Packages were being screened for explosives.

Walker: I do remember Ian Wilmut's personal assistant, Jackie, getting phone calls after it all hit the press. She had lots of phone calls, some of them were a bit crackpot, from people wanting their dogs cloned. The sadder ones were those people who had lost children or who had illnesses themselves, and this was going to be a breakthrough that could cure different diseases.

Colman: Dolly seemed to capture the imagination. It was a furry animal. Having a name that was identifiable helped enormously.

Bracken: If she'd been seen as being an animal that was locked away, that not many people saw, that could have perpetuated more bad publicity. But I think, because of the openness, that people were allowed to go and visit her and be shown around, this did help in the acceptance of the public.

Griffin: She performed well for camera, and everybody could see she was a perfectly normal animal. Because she was accessible and photogenic, she became the most famous sheep in the world. Any marketing manager would have killed for it. In some of the pictures it's as if she's interviewing the media.

Walker: I took a photographer down to see Dolly. This guy produced a kid's party crown, a little gold thing. I said: “I don't think we should.” We were all very keen not to allow Dolly to become humanized. She was a sheep and that was it.

Bracken: Away from the media and the cameras, we tried to treat her just like the other sheep, not as a sort of celebrity, which she obviously became.

Walker: The first time she was shorn, they took the wool — which I have some of, actually — to be knitted into a jumper for a cystic-fibrosis charity. Have you seen her in the museum? She's behind a glass case now because people kept pinching bits of wool from her. At least I got my wool while she was still alive.

Dolly lived for six and a half years and gave birth to several lambs herself. But in 2003, she began to show signs of illness.

Bracken: It was Valentine's Day. I think it was a Friday. We knew that there was the potential for this lung disease to have developed.

Griffin: She suffered from a disease called jaagsiekte. It's a disease of the lungs and one or two other sheep beforehand had gone down with it.

Wilmut: They thought she should be X-rayed over at the vet school. They were surprised at the size of the tumour in her lungs. We debated, under these circumstances, how hard we should struggle for her to recover. Wouldn't it be kinder to just let her go? So we euthanized her. You are responsible for the welfare of the animals on your project.

A decade later, another loss struck the scientific team with the death of Keith Campbell.

Colman: Keith was the driving force. He was the person who did the important experimental work that sowed the seeds of the protocol we all used. Dolly would not have happened without Keith.

Ritchie: Keith was, I suppose, 'unusual' is probably the thing you would say about him. He was quite hippy. He drove a Volkswagen Beetle, smoked roll-ups, had long hair.

Colman: He didn't have a great relationship with Ian. They were very different personalities and often argued.

Wilmut: I don't remember rows. We would have had slightly different priorities sometimes.

It's always very difficult to divide recognition up. What was obviously the cause of some annoyance and some criticism is that he didn't get the first authorship on the Dolly paper. He did get absolutely all the others. There was a time when he said the Megan and Morag paper was actually more important than Dolly. He definitely was frustrated that I got an FRS

(Fellow of the Royal Society) and ultimately a knighthood.

After a domestic dispute, Campbell killed himself on 5 October 2012.

Colman: Keith was a very good friend of mine and we used to go mountain biking in Scotland



“It literally was the cupboard.”

in the evenings after work. I spoke with him three days before he died. I was very shocked.

Walker: That hit me very hard, harder than I would have imagined. I hadn't seen him in many, many years. We were such a close, tight group at the time. We had to be.

Colman: I went to a meeting in Paris last January, where they had a posthumous award. They took a straw poll of how many people in the audience had been helped by what Keith had done, and a huge number of people put their hands up.

The techniques developed in the creation of Dolly were used to copy valuable livestock and make transgenic animals. But in biomedical labs, Dolly hinted at a future in which cells could be reprogrammed to an embryo-like state and used to treat human diseases.

Wilmut: The birth of Dolly turned the rules of development upside down, and made a lot of biologists think differently.

Jeanne Loring, stem-cell biologist, the

Scripps Research Institute, La Jolla, California: That was the onset of cattle cloning, which is actually quite popular now. There's a tremendous value in being able to improve cattle, and this gave people another tool.

George Seidel, animal reproductive biologist, Colorado State University, Fort Collins: There are cloned bulls producing semen that's being sold. There's an Angus bull called Final Answer, he's got half a million offspring or something like that. So his clone is called Final Answer II, and you can buy his semen at half the price. My wife and I have a cattle ranch, so we use Final Answer II. Hell, it's the same genetics. But from a theoretical standpoint, the transgenic stuff is really much more important than just making copies. To make our first transgenic cow, we created thousands of embryos. It was a huge effort. A tenth of the money, a tenth the animals is what transgenics plus cloning could do for you.

Robert Lanza, chief scientific officer, Astellas Institute for Regenerative Medicine, Marlborough, Massachusetts:

I was excited. Now we could hopefully apply the same technique — not so much for animals and agriculture — but for treating a long list of human diseases. What Dolly showed was the enormous power of that technology and the magic of the egg. There were factors in the egg that could take adult cells backwards in time and restore them to an embryonic state.

Shinya Yamanaka, stem-cell scientist, Kyoto University, Japan: My initial response was “Wow! It's like science fiction.” But it was not something I was planning to work on. Judging from the paper, the cloning process is very technically challenging. The next year, the first human embryonic-stem-cell paper came out. That's when I re-evaluated Dolly. I thought, at least in theory, we should be able to reprogram somatic cells back into the embryonic state so we can make ES-like stem cells directly from skin or blood cells.

McWhir: A result like Dolly stops people in their tracks, and they say: “Well hang on. If I'd have said that is impossible, what else am I saying is impossible?”

Schnieke: You have some experiments where it brings up your heartbeat. Dolly was one.

Ritchie: It's kind of like having children. I haven't got any myself. Maybe Dolly's that sort of child.

Wilmut: It would be wrong to say my name's known all the way around the world — but Dolly's is. ■

Ewen Callaway writes for Nature from London.



MYSTERY IN THE HEAVENS

*Ultra-powerful signals known as fast radio bursts are bombarding Earth.
But where are they coming from?*

BY ELIZABETH GIBNEY

No astronomer had ever seen anything like it. No theorist had predicted it. Yet there it was — a 5-millisecond radio burst that had arrived on 24 August 2001 from an unknown source seemingly billions of light years away.

“It was so bright, we couldn’t just dismiss it,” says Duncan Lorimer, who co-discovered the signal¹ in 2007 while working on archived data from the Parkes radio telescope in New South Wales, Australia. “But we didn’t really know what to do with it.”

Such fleeting radio bursts usually came from pulsars — furiously rotating neutron stars whose radiation sweeps by Earth with the regularity of a lighthouse beam. But Lorimer,

an astrophysicist at West Virginia University in Morgantown, saw this object erupt only once, and with more power than any known pulsar.

He began to realize the significance of the discovery¹ only after carefully going over the data with his former adviser, Matthew Bailes, an astrophysicist at Swinburne University of Technology in Melbourne, Australia. If the source was really as far away as it seemed, it had released the energy of 500 million Suns in just a few milliseconds. “We became convinced it was something quite remarkable,” he says.

But when no more bursts appeared, initial excitement turned to doubt. Radio astronomers have learnt to be sceptical of mysterious spikes in their detectors: the events can all too

easily result from mobile-phone signals, stray radar probes, strange weather phenomena and instrumental glitches. Wider acceptance of what is now known as the Lorimer burst came only in the past few years, after observers working at Parkes and other telescopes spotted similar signals. Today, the 2001 event is recognized as the first in a new and exceedingly peculiar class of sources known as fast radio bursts (FRBs) — one of the most perplexing mysteries in astronomy.

Whatever these objects are, recent observations suggest that they are common, with one flashing in the sky as often as every 10 seconds². Yet they still defy explanation. Theorists have proposed sources such as evaporating

WAYNE ENGLAND

The Parkes telescope in Australia detected the first fast radio burst in 2001.

black holes, colliding neutron stars and enormous magnetic eruptions. But even the best model fails to account for all the observations, says Edo Berger, an astronomer at Harvard University in Cambridge, Massachusetts, who describes the situation as “a lot of swirling confusion”.

Clarity may come soon, however. Telescopes around the world are being adapted to look for the mysterious bursts. One of them, the Canadian Hydrogen Intensity Mapping Experiment (CHIME) near Penticton in British Columbia, should see as many as a dozen FRBs per day when it comes online by the end of 2017.

“This area is set to explode,” says Bailes.

CURIOUSER AND CURIOUSER

Astronomers might have had more confidence in the Lorimer burst initially had it not been for a discovery in 2010 by Sarah Burke-Spolaor, who was then finishing her astrophysics PhD at Swinburne. Burke-Spolaor, now an astronomer at the US National Radio Astronomy Observatory in Socorro, New Mexico, was trawling through old Parkes data in search of more bursts when she turned up 16 signals that shook everyone’s confidence in the original³.

In most ways, these signals looked remarkably similar to the Lorimer event. They, too, showed ‘dispersion’, meaning that high-frequency waves appeared in the detectors a few hundred milliseconds before the low-frequency ones. This dispersion effect was the most important piece of evidence convincing Lorimer and Bailes that the original burst came from well beyond our Galaxy. Interstellar electrons in clouds of ionized gas are known to interact more with low-frequency waves than with high-frequency ones, which delays the low-frequency waves’ arrival at Earth ever so slightly, and stretches the signal (see ‘Flight delays’). The delay in the Lorimer burst was so extensive that the wave had to have travelled through a lot of matter — much more than is in our Galaxy.

Unfortunately for Lorimer and Bailes’ peace of mind, Burke-Spolaor’s signals also showed a crucial difference from the original: they seemed to pour in from everywhere, not just from where the telescope was pointing. Dubbed perytons, after a mythical winged creature that casts a human shadow, these bursts could have been caused by lightning, or some human-made source. But they were not extraterrestrial.

Lorimer decided to postpone his research into FRBs for a while. “I didn’t yet have tenure,” he says, “so I had to go back and do more mainstream projects, just to keep my research moving.” Bailes and his team kept going, and upgraded the Parkes detector’s time and frequency resolution. In 2013, they turned up four new FRB candidates that resembled the Lorimer burst⁴. But some outsiders remained sceptical that the signals were really coming from

space — not least because all the FRBs thus far had been seen by one team using one telescope. “I was desperate for someone else to find them somewhere else,” says Bailes.

In 2014, his wish was finally granted. A team led by astronomer Laura Spitler at the Max Planck Institute for Radio Astronomy in Bonn, Germany, published their observations of a burst at the Arecibo Observatory in Puerto Rico⁵. “I was ridiculously overjoyed,” says Bailes.

The Arecibo discovery convinced most people that FRBs were the real deal, says Emily Petroff, who is now an astrophysicist at the Netherlands Institute for Radio Astronomy in Dwingeloo. Yet, as long as the Burke-Spolaor signals went unexplained, they cast a shadow of doubt. “At any talks I would give,” says Petroff, “someone would say, ‘But what about

“THERE’S NO WAY THAT’S A MICROWAVE OVEN.”

perytons?” So in 2015, while still a graduate student at Swinburne, she led a hunt to track down the source of perytons once and for all.

First, Petroff and her team used the upgraded Parkes detector to pinpoint when the bursts were happening: at lunchtime. “Immediately I thought, ‘This isn’t weather,’” says Petroff. Then came another peryton at a suspiciously familiar radio frequency, which led the team to run experiments in the staff kitchen. Perytons, they discovered, were the result of scientists opening the microwave oven mid-flow. But the Lorimer event was in the clear: records showed that at the time of the burst, the telescope had been pointed in a direction that would have blocked any microwave signal from the kitchen⁶.

“So then I worried, maybe they’ve just got a different brand of microwave at Arecibo,” says Bailes, whose team at Parkes had, by then, racked up 14 separate bursts. He did not relax completely until later in 2015, when a burst was spotted at a third facility — the Green Bank Telescope in West Virginia. That burst had another quality that supported an extraterrestrial origin: its waves were rotated in a spiral pattern — which results from passing through a magnetic field — and were scattered as if they had emerged from a dense medium. “There’s no way that’s a microwave oven,” Bailes told himself.

BURSTS OF INSPIRATION

But that still leaves the question of what the FRBs actually are. The extreme brevity of the signal, just 5 milliseconds, implied that the source must be a compact object no more than a few hundred kilometres across — a stellar-mass

black hole, perhaps, or a neutron star, the compact core left over by a supernova. And the fact that Earth-based telescopes can detect the FRBs at all means that this compact source somehow puts out an immense amount of energy. But that still leaves a long list of candidates, from merging black holes to flares on magnetars: rare neutron stars with fields hundreds of millions of billions of times stronger than the Sun’s.

An important clue arrived earlier this year when Spitler’s team reported that at least one FRB source repeats: data from Arecibo revealed a flurry of bursts over two months, some spaced just minutes apart⁷. That behaviour has been confirmed by the Green Bank telescope, which detects signals in a different frequency band⁸. Until then, each of the observed FRBs had been a one-off event, which hinted at cataclysmic explosions, or collisions in which the sources were destroyed. But a repeating FRB implies the existence of a source that survives the pulse event, says Petroff. And for that reason, she says, “I would assume it would be something to do with a neutron star” — one of the few known objects that can emit a pulse without self-destructing.

Spitler agrees. As an example, she points to the Crab nebula: the result of a supernova explosion that was observed from Earth in 1054 and left behind a rapidly spinning pulsar surrounded by glowing gas. The Crab pulsar occasionally releases extremely bright and narrow radio flares, Spitler says. And if this nebula were in a distant galaxy and hugely boosted in energy, its emissions would look like FRBs.

If one source repeats, Spitler says, the simplest interpretation is that they all do, but that other telescopes haven’t been sensitive enough — or lucky enough — to see the additional signals. Yet others think that perhaps only some are repeating. “I wouldn’t be surprised if we end up with two or three populations,” says Petroff.

A LONG WAY HOME

Another crucial question is how far away the FRBs are. The 20 bursts seen so far seem to be scattered randomly around the sky rather than being concentrated in the plane of the Galaxy, which suggests that their sources lie beyond the borders of the Milky Way.

And yet to Avi Loeb, a physicist at Harvard University, such vast distances imply an implausibly large energy output. “If you want the burst to repeat, you won’t be able to destroy the source — therefore, it cannot release too much energy,” he says. “That puts a limit on how far away you can put it.” Perhaps, he says, the FRB sources are neutron stars in our own Galaxy, and the dispersion is mostly the result of still unknown electron clouds that blanket them.

But others suggest that such a dense cloud in the Galaxy should be visible in other wavelengths. At the California Institute of Technology (Caltech) in Pasadena, astrophysicist Shri Kulkarni has scoured data from several telescopes for a galactic source and turned up

nothing⁹. Kulkarni, who directs Caltech's optical observatories, initially argued for galactic FRBs, and even made a US\$1,000 bet on it with astronomer Paul Groot of Radboud University Nijmegen in the Netherlands. Now, he finds the evidence for extragalactic FRBs to be overwhelming, and has agreed to settle the bet — grudgingly. “I think I will pay him in \$1 bills,” he says.

Still, Kulkarni hasn't ruled out the possibility that the FRB sources lie in galaxies that are perhaps a billion light years away, rather than many billions. Such a distance would still require at least some of the signal dispersion to come from electron clouds in the host galaxy, he says. But closer FRBs would not have to be so energetic. “It takes them from being amazingly exotic, to just exotic,” he says.

The answer could mean a great deal to observers. If the FRB signals have travelled through local plasma clouds, they could give weather reports from neighbouring galaxies. But if they are truly cosmological — coming from halfway across the visible Universe — they could solve a long-standing cosmic mystery.

For decades, astronomers have known from observations of the early Universe that the cosmos should contain more everyday matter — the kind made up of electrons, protons and neutrons — than exists in the visible stars and galaxies. They suspect that it lies in the cold intergalactic medium, where it is effectively invisible. But now, for the first time, the dispersion of the FRB signals could enable them to measure the medium's density in any given direction. “Then, we have essentially a surgical device to do intergalactic tomography,” says Kulkarni.

RAPID-FIRE DETECTION

First, however, astronomers have to find a lot more FRBs and pin down their locations. “Until then, we are stumbling in the dark,” says Berger.

One way to accomplish that is to extract the FRBs from radio-telescope data in real time, so that scientists at other observatories can observe the bursts in multiple wavelengths. Since last year, the Parkes team has been doing this by boosting the observatory's in-house computing power, and scientists at Arecibo hope to follow suit this year. In February, the strategy seemed to be paying off when an independent team followed up within two hours of an FRB's detection at Parkes. The team tentatively pinpointed the burst to a specific galaxy almost 6 billion light years away. Further observations cast doubt on that interpretation. But even so, says Lorimer, the method is sound and may pay off in the future.

Others observers are putting their hopes in new telescopes. In 2014, astrophysicist Victoria

FLIGHT DELAYS

Astronomers are not sure what causes fast radio bursts (FRBs). But as the waves reach Earth, low-frequency ones lag behind high-frequency ones. The extent of this delay suggests that the signals have travelled through intergalactic space for potentially billions of light years.

SOURCE

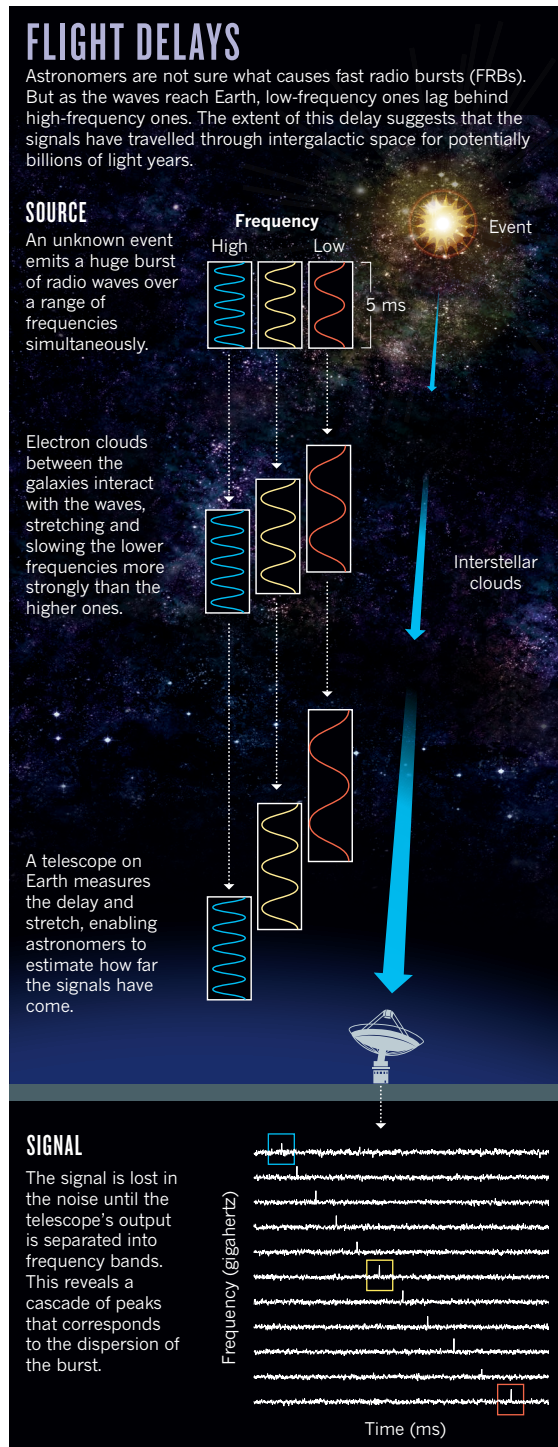
An unknown event emits a huge burst of radio waves over a range of frequencies simultaneously.

Electron clouds between the galaxies interact with the waves, stretching and slowing the lower frequencies more strongly than the higher ones.

A telescope on Earth measures the delay and stretch, enabling astronomers to estimate how far the signals have come.

SIGNAL

The signal is lost in the noise until the telescope's output is separated into frequency bands. This reveals a cascade of peaks that corresponds to the dispersion of the burst.



next year, says Kaspi, ultimately finding more than a dozen per day.

In Hoskinstown, Australia, meanwhile, Bailes and his colleagues are refurbishing the 1960s-vintage Molonglo Observatory Synthesis Telescope, turning it into an FRB observatory with a single half-pipe 16 times longer than CHIME's, although one-quarter as wide. The team has already found three as-yet-unpublished FRBs with the facility working at only about 20% of its final sensitivity, says Bailes.

Another strategy for locating the FRB sources is to work with existing facilities such as the Very Large Array: an 'interferometer' that uses the time difference between signals from 27 radio telescopes spaced across 36 kilometres of grassland near Socorro, New Mexico, to create a single, high-resolution image. Sometime in the next year or so, says Lorimer, the array could detect an FRB and locate its home galaxy. “Ultimately, that could settle a lot of arguments and bets,” he says.

Kulkarni, meanwhile, is leading two projects. The first uses ten 5-metre-wide dishes in an array that can see and locate only super-bright FRBs, but that makes up for its low sensitivity by peering at a huge swathe of sky. The second takes the principle to the extreme, using 2 antennas spaced at observatories 450 kilometres apart that will see only the very brightest FRBs, but that are able to examine half the sky at once. That would enable it to catch the rare FRBs that presumably exist within our own Galaxy, but whose extreme brightness existing telescopes are not designed to see. “Most facilities would just discount it as interference,” says Kulkarni.

If FRBs do turn out to come from cosmological distances, says Loeb, their identification would be a major breakthrough, potentially unravelling a new class of source that could be used to probe the Universe's missing matter. But then, he says, FRBs could also be something that no one has thought of yet: “Nature is much more imaginative than we are.” ■

Elizabeth Gibney is a reporter for Nature in London.

1. Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J. & Crawford, F. *Science* **318**, 777–780 (2007).
2. Champion, D. J. et al. *Mon. Not. R. Astron. Soc. Lett.* **460**, L30–L34 (2016).
3. Burke-Spolaor, S., Bailes, M., Ekers, R., Macquart, J.-P. & Crawford, F. III *Astrophys. J.* **727**, 18 (2011).
4. Thornton, D. et al. *Science* **341**, 53–56 (2013).
5. Spitler, L. G. et al. *Astrophys. J.* **790**, 101 (2014).
6. Petroff, E. et al. *Mon. Not. R. Astron. Soc.* **451**, 3933–3940 (2015).
7. Spitler, L. G. et al. *Nature* **531**, 202–205 (2016).
8. Scholz, P. et al. Preprint at <http://arxiv.org/abs/1603.08880> (2016).
9. Kulkarni, S. R., Ofek, E. O. & Neill, J. D. Preprint at <http://arxiv.org/abs/1511.09137> (2015).

COMMENT

POLICY Progress report from international panel of 300 social scientists **p.616**

MILITARY Anthropologist asks whether drones rewrite the rules of warfare **p.618**

HISTORY Museum refit showcases Scotland, workshop of the world **p.620**



BIODIVERSITY Tally of cryptic species changes figures dramatically **p.621**

KADIR VAN LOHUIZEN/NOOR/EYEVINE



An islander adds to coastal protection at Anse Kerlan beach in the Seychelles.

Make climate-change assessments more relevant

Stéphane Hallegatte, Katharine J. Mach and colleagues urge researchers to gear their studies, and the way they present their results, to the needs of policymakers.

With the ink just dry on the Paris climate agreement, policymakers want to know how they can act most effectively. Ambition is high: the long-term goal is to keep the average warming of the planet to well below 2°C, and even to 1.5°C. Governments, corporations and communities have many options for minimizing

dangerous climate change, and must choose between conflicting priorities and objectives. For example, how should governments decarbonize energy while increasing access to it without resorting to fossil fuels?

No single approach will work for all. The risks and impacts of climate change differ by place and time. Local values and

contexts matter. Small islands are vulnerable to sea-level rise, for example, and fossil-fuel exporters will lose profits from the transition to low-carbon energy. We must consider value judgements, such as the relative importance of economic damage versus biodiversity loss, as well as inequality and fairness.

And the relevant climate and social ►

► sciences are themselves diverse, from studies of the physics of storm formation to investigations of the role of heritage in cultural identity. The challenge for those who assess such scientific knowledge, such as the Intergovernmental Panel on Climate Change (IPCC), is to summarize results in ways that are true to the original research, explicit about the values and judgements in the analysis, and digestible by and useful to policymakers and the public.

For example, the IPCC's 2014 *Synthesis Report* encapsulated factors from climate to ecology to technology into a single figure (Figure SPM.10)¹. This illustrates how long-term global risks are linked to emissions-reduction requirements under different physical, policy and risk scenarios. Such a figure, although an achievement, can convey only a glimpse of the complex analysis that went into it.

In the IPCC's sixth cycle of assessment, the climate-science community needs to supply the right sorts of information to help decision-makers to construct policies from myriad mitigation and adaptation options. Producing this information will require more multidisciplinary research, updated strategies for communicating uncertainty and studies of a broader range of climate and risk projections that include the impacts of policy responses.

Here, we set out four steps to putting policy relevance at the core of both research and assessment.

Integrate disciplines from the start. The range of risks summarized in the IPCC's 2014 *Synthesis Report* was limited by the research available. For example, the assessment highlighted increasing risks of climate extremes but said little about how climatic hazards interact with societal vulnerability. Sparse information on how risks evolve at specific warming levels resulted in the reporting of broad, qualitative levels of risk — for example, 'undetectable' to 'very high', as judged by experts. But comparison across risks was difficult.

Climate scientists need to close these gaps by scrutinizing the feedbacks between development pathways, climate change and its impacts and risks, and policies and responses. The community has created socio-economic scenarios that are better able to combine climate-policy consequences and climate-change impacts in certain areas — such as how poverty reduction reduces vulnerability to extreme events — and to investigate their interplay with development trends ranging from population to land-use trajectories². But covering many climatic and societal futures, globally to locally, is a monumental task. Projects that compare assumptions and results between different models are a start, but need to include more



Sugarcane production is rising in Brazil to meet demand for ethanol for biofuel.

evidence and expert judgements across disciplines.

Research and assessments must be designed to solicit and answer questions crucial to decision-making. For example, how do risks and requirements compare for a climate goal at 1.5 °C, 2 °C or more? How can we avoid locking in to carbon-intensive development pathways and keep open options for rapid decarbonization? How can the effectiveness of adaptation actions be ensured? And how can emissions be reduced without slowing the pace of poverty reduction?

Explore multiple dimensions. Risks from a changing climate and responses to it vary dramatically from place to place, through time and with different levels of adaptation and mitigation. Projections of increases in sea level for different emissions scenarios, for example, range from tens of centimetres to more than 10 metres over centuries to millennia³. Small islands might quickly face inundation whereas large countries would have more time to adapt. Past assessments focused on characterizing a few alternative futures (such as continued high emissions versus ambitious mitigation) rather than weighing up the risks and benefits of limiting warming across a ladder of possible targets: 1.5 °C, 2 °C, 2.5 °C or higher.

A broader census of differences through space and time would strengthen the information foundation for policymaking.

Decision-makers with different goals could select portfolios of responses, for example, based on risks to all, risks to the most vulnerable, risks of economic damages, risks of irreversible changes or a combination.

The distribution of losers and winners — regarding policies and impacts as well as people and places — needs to be studied. For example, the destruction of coral reefs affects fishing communities and may add to stresses, especially in places with weak governance. In some high-latitude areas, by contrast, a warming climate will bolster agricultural yields. Building sea walls could reduce coastal flood risks but threaten ecosystems, historical heritage and landscape beauty. Risks and opportunities from investments in mitigation options need to be evaluated. For example, expanding biomass energy may reduce (or reverse) emissions but could also threaten food production and biodiversity. Renewable energy reduces emissions and provides electricity more cheaply than that from fossil fuels in many remote locations, where some of the poorest people live.

More research is needed on regional challenges and opportunities that go beyond the use of a single metric — global mean warming — as a proxy for climate change and its impacts⁴. For example, ocean acidification and sea-level rise are not linearly related to peak temperature, and the risks that they create require more detailed investigation. And reducing emissions of short-lived climate



pollutants such as soot and tropospheric ozone precursors might not change peak warming, but would slow the rate of warming globally³; this would allow more time for ecosystems and societies to adapt, as well as provide local health benefits.

Consider uncertainty. Decision-makers need to appreciate a wide range of possible outcomes, including uncertainty in the consequences of global climate policies. Four aspects of uncertainty must be evaluated and communicated: probability ranges that can be narrowed with future research; unknowns that are linked to a deep lack of knowledge; uncertain reactions that depend on societal decisions and geopolitical events; and other areas of uncertainty that reflect random or chaotic features of the climate system.

The implications of these uncertainty types for policymaking and research need to be untangled. Those that relate to underlying Earth-system processes, such as climate mechanisms that we do and do not understand, or the inherent variability of the climate system, can be addressed through research that increases understanding of climatic hazards. Extreme events and resulting damages lie in the tails of probability distributions that are inherently difficult to quantify or even characterize qualitatively.

Uncertainty need not be a bad thing. Uncertainties related to human choices — such as the multiple pathways to achieve a climate goal — can offer flexibility⁶.

For example, much of the uncertainty in the relationship between emissions in 2050 and eventual temperature rise stems from the possibility of compensating for modest short-term emissions reductions with larger efforts, including negative emissions, in later decades.

An awareness of the diversity of options and their risks is important for making smart policies that allow for regular revisions in light of new information and feedback. More ambitious near-term emissions reductions create more flexibility for responses through the century, depending on whether useful and affordable technologies become available and how climate impacts pan out. Less mitigation early on would constrain options later and compound risks⁷. Short-term actions — such as the commitments for 2025 or 2030 that countries have made towards the Paris Agreement — can be compatible with a range of long-term targets, depending on the ambition of our efforts later in the century.

Assessing whether current policies are consistent with long-term goals depends on many factors that are impossible to predict with confidence^{8,9}. And not knowing how people will respond makes such an assessment even harder. So emissions pathways that seem compatible today with a long-term temperature target could lead us to higher — or lower — levels of warming, depending on everything from future global climate policies to technology costs to the climate sensitivity of the Earth system. Intensified focus on limiting global warming to 2 °C or 1.5 °C decreases the risk of greater warming in the long term, for example a rise exceeding 3 °C, should available technologies turn out to be limited or climate sensitivity higher than expected.

Researchers need to assess how different sources of uncertainty affect decision-making, especially in worst-case scenarios. What should we do if temperatures start to rise more rapidly or the impacts are more dangerous than we expect? How can we detect such departures and how should we alter course? Climate policies might prove to be harmful and need revising; technology costs might not fall; carbon capture and sequestration might not work.

Inform holistic solutions. A fuller evaluation of risks and options is needed that includes those created by climate-change responses for other policy goals. For example, the assessment of climate-change risks at 1.5 °C in the IPCC's 2014 *Synthesis Report* foresaw impacts on coral

reefs, Arctic sea ice, water availability, food production and sea-level rise. But the bigger picture should also include issues related to climate mitigation, such as economic duress, land- and water-use trade-offs and calls for high-risk geoengineering methods.

The impacts of climate changes and climate policies will interact if, for instance, a slower reduction in poverty owing to higher energy costs increases vulnerability. Synergies and trade-offs must be evaluated, including risks arising from mitigation actions — not just inactions. Social and climate scientists must investigate the political and socio-economic impacts of climate policies (short- as well as long-term), the distribution of those who benefit and those who are adversely affected, and the influences of powerful interest groups.

It is important to explore how climate responses can advance the Sustainable Development Goals and especially poverty reduction¹⁰. For instance, improving access to clean energy and decreasing the economic impacts of extreme weather events can accelerate development progress while protecting poorer nations against climate change. Climate action and protection will never be the sole priorities for decision-makers, but they will be integral to the full policy landscape. Research and assessment can create a powerful foundation for these interactions, and empower decisions in the years ahead. ■

Stéphane Hallegatte is senior economist in the Climate Change Policy Team at The World Bank, Washington DC, USA.

Katharine J. Mach is a senior research associate in the Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA.
e-mail: shallegatte@worldbank.org

1. Pachauri, R. K. et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2014).
2. O'Neill, B. C. et al. *Clim. Change* **122**, 387–400 (2014).
3. Clark, P. U. et al. *Nature Clim. Change* **6**, 360–369 (2016).
4. Steinacher, M., Joos, F. & Stocker, T. F. *Nature* **499**, 197–201 (2013).
5. Rogelj, J. et al. *Proc. Natl Acad. Sci. USA* **111**, 16325–16330 (2014).
6. Otto, F. E. L., Frame, D. J., Otto, A. & Allen, M. R. *Nature Clim. Change* **5**, 917–920 (2015).
7. Edenhofer, O. et al. (eds.) *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2014).
8. Kriegler, E. et al. *Technol. Forecast. Soc. Change* **90**, 1–7 (2015).
9. van Vuuren, D. P. & Riahi, K. *Clim. Change* **104**, 793–801 (2010).
10. Hallegatte, S. et al. *Shock Waves: Managing the Impacts of Climate Change on Poverty* (World Bank, 2016).

A list of co-signatories accompanies this article online at go.nature.com/28jfsrq.



JOHN STANMEYER/NGC

Migrant Somalis crowd the night shore of Djibouti trying to capture inexpensive mobile-phone signals.

Social-progress panel seeks public comment

Marc Fleurbaey and colleagues explain why and how 300 scholars in the social sciences and humanities are collaborating to synthesize knowledge for policymakers.

International panels of experts have proliferated to marshal scientific knowledge. There are panels on climate change, biodiversity, chemical pollution, food security and nuclear proliferation. All are concerned with long-term issues that have profound economic, social, political and cultural ramifications.

Those issues, and the uncertainties around them, represent unprecedented challenges for our societies. Many of the obstacles to the identification and implementation of solutions to the 'wicked problems' — those that are multifactorial and exceptionally complex — come from inertia and misalignment in institutions, conventions and forms of collective action¹. Meanwhile, we are still facing many classic threats — war, violence and terrorism produce major disruptions and instabilities while widening inequalities put increasing strains on social cohesion.

All this questions our collective capacity to deliver on global sustainability goals and to

ensure a viable future for subsequent generations. Responses, so far, do not live up to the urgent and critical nature of the challenges. Impressive efforts have been made around the adoption of the United Nations' Sustainable Development Goals and the Paris climate agreement. However, dysfunctional, short-sighted policymaking in many countries is profoundly worrisome, as is foundering cooperation in international arenas such as the World Trade Organization and the European Union, and the return of anti-democratic trends. The absence of a positive and cohesive long-term vision of what we could collectively aim for is one key factor responsible for this helplessness and impotence.

That vision is the mission of a new panel convened last year, the International Panel on Social Progress (IPSP). It comprises more than 300 social-science and humanities scholars coordinated by the Fondation Maison des Sciences de l'Homme in Paris and by Princeton University in New Jersey. The

IPSP is preparing a report on directions that could be taken in the twenty-first century to create better societies. We are members of the panel's steering committee, and two of us (R.K. and H.N.) are co-chairs of its scientific council. In the next few months, the IPSP will release the first draft of its report.

We call on researchers, policymakers, think tanks, companies, non-governmental organizations (NGOs) and citizens to provide us with feedback during the comment period. From August to December 2016, interested parties will be able to weigh in on the panel website, www.ipsp.org, which will host a comment platform, discussion forums and surveys. Informed by these views, we hope that the final report will reflect an open and broad international debate on 'mobilizing utopias'.

SYNTHESIS REPORT

Modern science and technology have been nurtured by a fervent belief that they lead to social progress. It has become clear that the

relationship is more complex. Considerable developments in the social sciences and the humanities since the Second World War have brought a much better understanding. For instance, the virtues and limitations of market economies and public interventions have been extensively scrutinized at the intersection of economics, political science, sociology and anthropology. The drivers of inequality and its possible remedies are still debated, but that discourse is now much more advanced, thanks in particular to better data.

These developments have coincided with growing specialization between and within disciplines, and an increasing awareness of the diversity of regional perspectives. This makes it impossible for a single scholar or even a small group of experts to synthesize the accumulated corpus of knowledge.

Creating such a synthesis that will be accessible to policymakers and social actors therefore requires a large, coordinated effort. The IPSP brings together scholars from economics, sociology, political science, law, anthropology, history, science and technology, and philosophy. The panel includes representatives from around the world, with about 40% of them women. The geographical composition of the panel is proportionally representative of national academic output, which unfortunately means that non-Western countries are under-represented relative to their populations, and many of the representatives they do have work in developed countries. The panel is split into 22 thematic chapters and five cross-cutting groups (science and technology, gender, migrations, health and social movements).

The chapters are grouped into three parts. One on socio-economic transformations examines economic growth and environmental constraints, inequalities, labour, urbanization, markets, corporations and the welfare state. The part on governance explores trends and options for democratic institutions, the rule of law, global governance, multinational organizations, violent conflicts, as well as the evolving forms of communications and media. The third part studies the sources and consequences of transformations in cultures and values, religions, families, health and education, as well as identities and social bonding.

Issues addressed include: the shift from production and consumption to well-being; the importance of urban design in shaping social relations; the transformation of the role of the welfare state; the questioning of democratic institutions in a globalized world and the contested diversification of family and sexuality.

BROAD INFLUENCE

This broad scope will make it possible, in the final report, to propose a systemic perspective on the evolution of societies in the world. Such a bird's-eye view has been largely left to

the media and popular pundits, and needs to be reclaimed by the scholarly community through the mobilization of its expertise.

Members of the IPSP hope to influence various audiences and processes. The main goal is to enter into a dialogue with citizens, social actors (NGOs and think tanks) and policy-makers, providing them with useful ideas that can enrich the debates, and guide actions.

Another goal is to reach researchers as well as local, national and international research organizations. The Intergovernmental Panel on Climate Change (IPCC), for instance, has galvanized research in climate science and policy through a similar contribution. The IPSP report will provide a critical review of the literature on social progress, identifying areas of consensus, controversial points and knowledge gaps. The effort will also propose innovative insights, including alternative policy narratives and ways to frame problems. For instance, the reduction of inequalities is usually discussed in terms of income tax and wealth redistribution, but can also be pursued by governance of the labour market and new revenue from environmental policy.

To reach this diverse audience, the IPSP will produce a variety of outputs. The three-volume report will bring together the work of all 22 chapters. A smaller book, written by a small team and for a general audience, will distil the main narrative and conclusions. We will also produce a policy toolkit of recommended actions for all types of actors, as well as video interviews and talks. This approach to diffusion emulates bodies such as the Organisation for Economic Co-operation and Development (OECD), which has, for instance, an interactive website for its Better Life Index (www.oecdbetterlifeindex.org).

To reach this diverse audience, the IPSP will produce a variety of outputs. The three-volume report will bring together the work of all 22 chapters. A smaller book, written by a small team and for a general audience, will distil the main narrative and conclusions. We will also produce a policy toolkit of recommended actions for all types of actors, as well as video interviews and talks. This approach to diffusion emulates bodies such as the Organisation for Economic Co-operation and Development (OECD), which has, for instance, an interactive website for its Better Life Index (www.oecdbetterlifeindex.org).

PROS AND CONS

The IPSP process may be useful to other panels. The report will make recommendations — in contrast to the IPCC, which avoids prescriptive language (although special efforts were made in its latest report to clarify the ethical issues and value judgments involved in policymaking²). To respect the diversity of views among readers and users, every IPSP recommendation will be associated with a clear acknowledgment of the underlying values and assumptions.

Can the panel have an impact on decision-makers? Unlike the IPCC, the IPSP is not a government-led institution. It is a bottom-up initiative coordinated by social scientists and scholars who are free to delineate the scope of their analysis. This independence has pros and cons.

On the plus side, the report can be more nimble and frank than those of the

IPCC, say, not having to be approved by governments. We hope that such freedom will be an advantage for reaching a larger set of actors — such as NGOs, trade unions, think tanks and activists. And the consultation exercise in the second half of this year will include meetings with decision-makers such as United Nations organizations, the World Social Forum, the World Economic Forum, the OECD and the World Bank.

Yet the IPCC-style official approval processes does provide an institutional space where experts and decision-makers air their perspectives³. Without this structure, there is a risk that the report won't meet the demand of potential users. Finally, the IPSP relies entirely on the goodwill of busy scholars and their willingness to make time for this collective endeavour without any financial compensation. Many of them contribute part of their institutional budget to cover some of their costs, a remarkable proof of commitment.

Can such a diverse group produce a strong message? The goal of laying out the 'state of the art' in contested domains is not unreasonable: existing panels tackle social and economic policy issues (for example, the IPCC working groups on climate adaptation and mitigation policies), and controversies also arise around more technical subjects such as nuclear proliferation or biodiversity⁴. At the IPSP, we consider that disagreement requires panellists to focus on objective reviews of ongoing debates rather than seeking consensus at the cost of substance and depth. Diversity of approaches is an asset rather than a liability.

The panel is itself an experiment in whether the social sciences can, in this format, make a difference in the quest for social progress. That requires not only academic input, but also a broad, open and lively public debate. Please have your say! ■

Marc Fleurbaey is professor in economics and humanistic studies at Princeton University, New Jersey, USA, and heads the economics of well-being programme at the Collège d'Études Mondiales-FMSH in Paris, France. **Olivier Bouin, Marie-Laure Djelic, Ravi Kanbur, Cécile Laborde, Helga Nowotny, Elisa Reis, Elke Weber, Michel Wieviorka and Xiaobo Zhang.** e-mail: mfleurba@princeton.edu

1. IPCC. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) (Cambridge Univ. Press, 2014).
2. Edenhofer O. & Minx, J. *Science* **345**, 37–38 (2014).
3. Dubash N., Fleurbaey, M. & Kartha, S. *Science* **345**, 36–37 (2014).
4. US National Academy of Sciences, Engineering, and Medicine. *Genetically Engineered Crops: Experiences and Prospects* (NASEM, 2016).

Full author details accompany this article online at go.nature.com/28l9n7u.



A US Predator drone in Kandahar, Afghanistan.

MILITARY TECHNOLOGY

Death by remote control

Ann Finkbeiner examines a study that probes how drones have ‘remixed’ warfare.

A drone is a good way to kill someone. A pilotless aeroplane watches from above, targets and shoots with precision, and disappears into the sky. The killer is never in danger. Drones can fly at 160–300 kilometres per hour at altitudes of up to 15,000 metres, hover over a target for hours, ‘see’ people and objects (clearly enough to read a car number plate from 3 kilometres up), and detect mobile-phone signals. At one-tenth to one-hundredth of the price of fighter planes, drones are practically disposable.

Little surprise, then, that by 2012, one-third of the US Air Force’s aircraft were drones, and half of the pilots in the Air Force were being trained to fly them. The US military has used drones for fighting in Afghanistan, Iraq, Yemen, Somalia, Libya, Pakistan and the Philippines. This has “remixed” warfare, anthropologist Hugh Gusterson argues in *Drone*, an overview of the implications and ethics. If war is a duel in which both sides are vulnerable, then drone warfare may not even be war.

➔ **NATURE.COM**
For more on science in culture see:
[nature.com/
booksandarts](http://nature.com/booksandarts)

It is so asymmetrical that it resembles hunting, he writes — a “new form of state violence”, harder to define and control with national and international laws. As such, he argues, “the drone is an inherently colonialist technology that makes it easier for the United States to engage in casualty-free and therefore debate-free intervention”.

Drone operation is bizarre. US pilots sit in grubby trailers in front of computers somewhere in the depths of the Nevada desert; Gusterson calls them “stick monkeys”. The screens show live videos of roads, compounds, people — views also seen by intelligence analysts, commanders and military lawyers, all of whom inform the decision to fire. Then a ‘sensor operator’ aims a laser at the person or vehicle targeted, the drone pilot triggers a missile that follows the laser,



Drone: Remote Control Warfare
HUGH GUSTERSON
MIT Press: 2016.

and 15–30 seconds later, they watch the infrared flash, the flames and the incineration of the enemy.

Some might imagine that this precision makes drones more humane compared to, say, the bloodbath of the Second World War’s Battle of the Bulge. But it does not prevent civilians from being killed: if someone wanders within screen view at the last second, or a targeted soldier passes his phone to a relative, it can happen. Those who live under circling drones — interviewed by foreign correspondents or aid workers — report chronic terrorization. And according to a study by the US Department of Defense, Gusterson recounts, half of the US pilots, despite being on the other side of the planet, have “high levels of operational stress” and can experience post-traumatic stress disorder. Targeted killing ultimately means only that fewer civilians die. Nobody knows how many: estimates depend entirely on who does the estimating.

The standard rules of warfare no longer apply. The Geneva Conventions and their protocols — established from 1949 to 2005

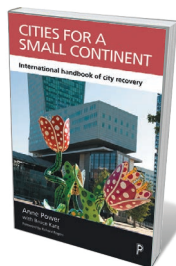
VERONIQUE DE VIGUERIE/GETTY

to protect civilians, among others — define battlefields and combatants. But insurgent wars have no clear battlefields and no full-time, uniformed fighters. Combatants are the people listed as targets for drone strikes. The list, Gusterson writes, is “maintained by U.S. military and intelligence agencies”, and includes people not known to be terrorists but possessing that profile. A person is added after analysts balance their importance against the number of civilians likely also to be killed. Gusterson wonders about the line between legal targeted killing and illegal assassination.

Although 76 countries use drones, Gusterson mentions only the United Kingdom, Israel, Iran and the United States, focusing on the last. He writes that he doesn't intend this book as an argument for or against drones and that he wants only a debate on regulating their use, yet he is clearly against them.

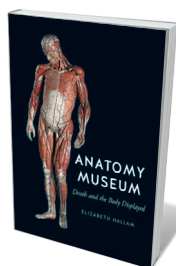
A reader trying to decide whether drones are a more humane form of weaponry, a hunting party, a neocolonial ploy or all three at once, will want to look closely at the author's sources. Anthropologists have traditionally gone into the field, meeting, observing and listening to their sources. Gusterson's sources are predominantly journalists, memoirists and authors who report

Books in brief



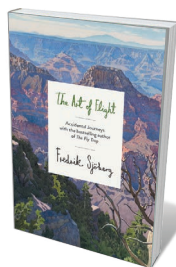
Cities for a Small Continent: International Handbook of City Recovery
Anne Power POLICY (2016)

Many of Europe's storied cities have seen more bust than boom for decades, writes urban-sustainability specialist Anne Power. Yet a number have risen reinvented, and in this brilliant analysis, Power shows how. She follows the march of seven “Phoenix cities” with strong industrial legacies, from Sheffield, UK, to Turin in Italy, as they weather upheavals and de-industrialize with the aid of major public investment. These conurbations should be seen, she argues, as the vanguard in the low-carbon transformation outlined by economist Nicholas Stern (M. Grubb *Nature* **520**, 614–615; 2015).



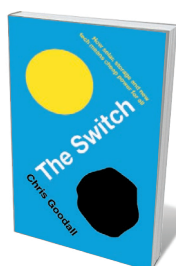
Anatomy Museum: Death and the Body Displayed
Elizabeth Hallam REAKTION (2016)

Pickled in formalin, stripped down to articulated skeletons or depicted in wax or plastic, human anatomical remains have educated generations of medics and fired the public imagination. Anthropologist Elizabeth Hallam uses the Anatomy Museum at the University of Aberdeen, UK, to anchor a history of such collections as “synoptic mazes” — labyrinthine summations of knowledge. Hallam charts their convoluted chronicles of acquisition, dissection and preservation, weaving in a narrative on the cultural display of death, from ancient ossuaries to plastinated bodies.



The Art of Flight

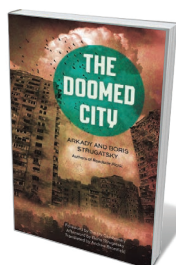
Fredrik Sjöberg (Translated by Peter Graves) PARTICULAR (2016)
Entomologist Fredrik Sjöberg's best-selling memoir *The Fly Trap* (Particular, 2014) marked him as a maestro of the episodic. Here, he completes a trilogy with two books in one — “accidental journeys” by fellow Swedes whose omnivorous curiosity rivalled his own. *The Art of Flight* focuses on Gunnar Widforss, exalted in the United States for his haunting landscape paintings of national parks. *The Raisin King* tackles polymath Gustav Eisen, who studied earthworms, *Anopheles* mosquitoes and viticulture, brought avocados to California and sparked the founding of Sequoia National Park. A joy.



The Switch

Chris Goodall PROFILE (2016)

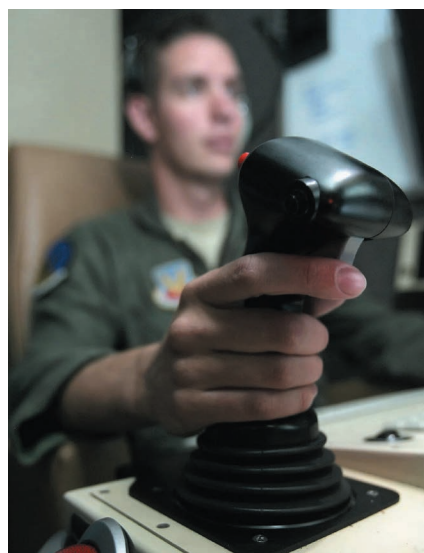
The world is poised for the solar revolution, argues energy writer Chris Goodall in this nippy, number-crunched study of the science behind the “switch”. Noting that solar farms will have to cover 1% of Earth's surface by 2050 to meet global energy needs, he treads the road towards that goal. He examines readiness in industry and banking, research on new solar-collection materials such as perovskites, the state of back-up renewables and innovative batteries. With many governments and some utility companies primed for action, Goodall avers, the fossil century could be history within two decades.



The Doomed City

Arkady Strugatsky and Boris Strugatsky (Translated by Andrew Bromfield) CHICAGO REVIEW PRESS (2016)

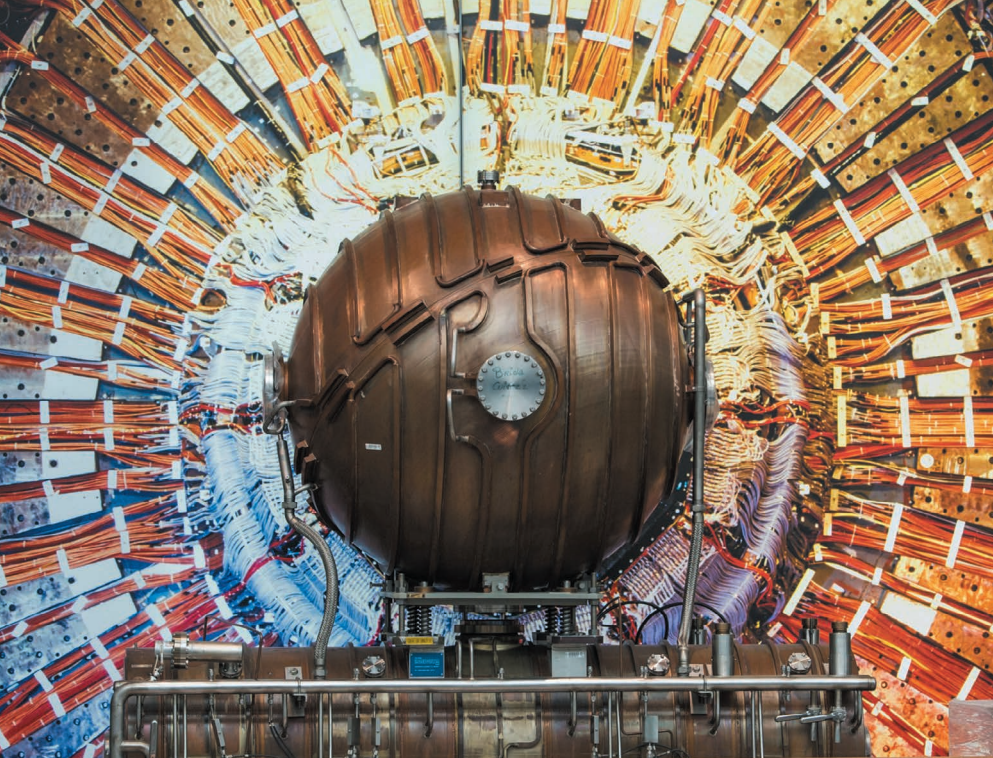
Doyens of Russian science fiction Arkady and Boris Strugatsky wrote this nihilistic ‘lost’ novel in the 1970s. In its English-language debut, we are dumped abruptly into the Experiment, a garbage-choked, baboon-infested city with an artificial sun and an eerily mismatched populace. Here, astronomer-turned-rubbish-collector Andrei begins a grim trek into the ideology of tyranny. A book that carries an Orwellian punch, and a crazed energy all its own. [Barbara Kiser](#)



A drone operator in Nevada.

regularly on the field or are veterans of it. Some anthropologists do analyse a culture remotely, on the basis of others' experiences and impressions. As a journalist, I would rather do my own reporting and not hover drone-like over the field that others have defined. ■

Ann Finkbeiner is a freelance science writer in Baltimore, Maryland, and author of *The Jasons*. She blogs at *The Last Word on Nothing*, www.lastwordonnothing.com. e-mail: anniefk@gmail.com



Part of CERN's Large Electron-Positron Collider, on display at the National Museum of Scotland.

MUSEUMS

Workshop of the world

Colin Macilwain talks to the curators of the National Museum of Scotland on the eve of a grand expansion.

The afterlife of Dolly the sheep will take an exciting turn next month: the clone is the first attraction of a new, permanent exhibition at the National Museum of Scotland (NMS). Dolly, who died in 2003 and survives in taxidermied splendour, will greet visitors at the entrance to a grand atrium housing ten refurbished science, technology and design galleries in the Victorian building, which sits beside the University of Edinburgh.

Since Dolly's leap to fame (see go.nature.com/1ujdd4k), the anticipated commercial bonanza in cloning has not occurred. Nevertheless, the bovid is part of a rip-roaring story of how a nation of 5 million people helped to forge the modern age.

"Scotland was the workshop of the world," explains Klaus Stauber, a science historian and principal curator of technology at the NMS. He alludes to the turn of the last century, when Glasgow's Clydeside hosted as much as one-quarter of global shipbuilding and locomotive production. But he brushes off any suggestion that the transformation from powerhouse to tourist attraction is a metaphor for Scotland's recent history that runs too close to the bone. "The Scots have always had this ability to reinvent themselves," Stauber says, noting the 'Silicon Glen' of the 1980s. Here, much of Europe's computer hardware was assembled. More recent, if scattered, successes have emerged

in computer gaming (*Grand Theft Auto* originated here) and biomedical science.

The range of exhibits being installed does justice to Stauber's grand narrative. There is an accelerating ring from the disused Large Electron-Positron Collider at CERN (Peter Higgs, namesake of the boson, is at the University of Edinburgh). There's the story of Glasgow pharmacologist and Nobel laureate James Black, who had a central role in developing beta blockers. And there will be a stunning array of electrical and mechanical engineering, from the world's oldest Stirling engine to the tip of a modern wind-turbine blade.

The unifying theme of the six new science and technology galleries is to follow scientific concepts through to technology and production. The museum's wider collection reflects diligent hoarding of bits and pieces from the Industrial Revolution. "Many Scots have held on to their working machinery," says Stauber. His research

A model of a heat engine made by Robert Stirling in 1816.

National Museum of Scotland gallery opening
8 July 2016.
Edinburgh, UK.



includes the history of the machine-tool industry, due to feature in a manufacturing hall. (Europe's first numerically controlled machine tool was built by electronics firm Ferranti in Edinburgh in the 1960s.)

The NMS is Britain's most popular museum outside London, with 1.6 million visitors last year. It has enjoyed steady public and private support, while many museums globally are suffering from cuts. The £14-million (US\$20-million) refurbishment is supported by the UK national lottery, biomedical charity the Wellcome Trust, the Scottish government and private donors. It is part of a £80-million, 15-year 'masterplan' to transform the premises, built in 1866.

Touring the exhibition areas last month, surrounded by workers assembling displays, I saw how technology is transforming the craft of exhibiting itself. For example, layers of interactivity allow visitors to delve as deep as they like. But there will also be older-school approaches on display. "Whenever I tell a taxi driver what I'm doing, the first thing they want to know is if there'll be plenty of buttons to push," says Stauber.

The museum's first director, George Wilson, was a professor of technology at the University of Edinburgh, and the institutions remain closely linked. Most museum curators teach at the university, and many are involved in joint research projects.

Right now, the priority is the logistics of installing some 3,000 objects — three-quarters on public display for the first time — in time for the 8 July opening. Elsa Cox, curator of the energy hall, Energise, has just collected a control console from the decommissioned Murchison oil platform in the North Sea. One of the challenges, she says, is getting industrial contributors to take her deadlines seriously.

Energise exhibits will include an early prototype of Salter's duck — a key device in the history of wave power — and the switch that connected the now-defunct Dounreay fast breeder reactor in Caithness to the grid. Visitors will get the chance to power a city for a day, Cox says, learning about the trade-offs between different sources of energy.

The display won't take sides on the heated argument between the Scottish government, which backs wind-power, and the government of Britain at large, which has cut support for renewables and wants to increase nuclear capacity. "The Scottish government funds us," says museum spokesman Bruce Blacklaw. "It doesn't govern us." ■

Colin Macilwain is a science-policy journalist, based in Edinburgh, and editor of the policy newsletter Research Europe.
e-mail: cfmworldview@googlemail.com

Correspondence

Count cryptic species in biodiversity tally

The race to describe and archive the planet's dwindling biodiversity (see K.-D. B. Dijkstra *Nature* **533**, 172–174; 2016) becomes even more urgent with the realization that the task's scale may be an order of magnitude greater than estimated.

Dijkstra notes that we have so far named only about 1.2 million of Earth's estimated 8.7 million or so eukaryotic species. Such estimates are based largely on counts of invertebrate 'species' that are visually distinguishable ('morphospecies'). However, genetic analysis has revealed that many supposedly uniform morphospecies are complexes of multiple, reproductively isolated lineages, each of which constitutes a separate but cryptic species (D. Bickford *et al. Trends Ecol. Evol.* **22**, 148–155; 2007).

These discoveries boost the biodiversity of even the largest vertebrates, such as elephants. The effect is greater in small vertebrates (such as lizards and frogs) and in invertebrates, which are often complexes of ten or even more species (P. M. Oliver *et al. BMC Evol. Biol.* **10**, 386; 2010). The quoted estimate of the number of (morpho)species on Earth could therefore be just 10% of the true species number.

Michael S. Y. Lee *South Australian Museum and Flinders University, Adelaide, Australia.*

Paul M. Oliver *Australian National University, Canberra.*
mike.lee@samuseum.sa.gov.au

Fewer papers would scotch early careers

Daniel Sarewitz argues that the pressure to publish is fuelling irreproducibility, but we disagree that the solution is to publish fewer papers (*Nature* **533**, 147; 2016). In today's competitive arena, asking this of scientists — particularly junior ones — is to ask them to fall on their swords.

Investing more effort in fewer

but 'more complete' publications could hold back early-career researchers, who already face fierce competition. To generate a first-author publication, graduate students on average take more than a year longer than they did in the 1980s (R. D. Vale *Proc. Natl Acad. Sci. USA* **112**, 13439–13446; 2015). Introducing further delays for junior scientists is not an option as long as performance is rated by publication metrics.

In our view, publishing less is not a feasible or responsible way to improve data quality. This would be better achieved by increasing the transparency of peer review and by introducing alternative metrics as indicators of reproducibility. Science's goal is to share as much information as possible — not to withhold it.

Gary S. McDowell, Jessica K. Polka *The Future of Research, Abington, Massachusetts, USA.*
jessica.polka@gmail.com

Pre-emptive action against EU invasives

As appointed representatives of the European Union's Scientific Forum on Invasive Alien Species, we wish to point out that the EU regulation states priority should be given to the listing of invasive species "that are not yet present in the Union or are at an early stage of invasion" (see M. Lehtiniemi *et al. Nature* **533**, 321; 2016).

The risk assessments needed to include species on the EU list of concern were already available for most of the 37 species (see go.nature.com/gigftz) and were a natural starting point. Moreover, listing is a political process: actions to protect biodiversity are weighed against factors such as socio-economic interests. Adequate evidence is necessary to ensure that proper action is taken.

Notably, none of the targeted 37 species is established in all EU member states, and all have potential for future spread (see go.nature.com/28vtjpk).

Any member state, including those where Lehtiniemi and

colleagues are based, can develop and submit risk assessments for candidate species. The scientific forum assesses these according to regulation criteria. Member states and the European Commission then decide whether to regulate species at the EU level.

Johan Näslund *Swedish Environmental Protection Agency, Stockholm, Sweden.*
Erland Lettevall *Swedish Agency for Marine and Water Management, Gothenburg, Sweden.*
johan.naslund@naturvardsverket.se

Supporting women postdocs in Israel

Israel's Weizmann Institute of Science is ranked among the world's top research institutions, but only 16.5% of its faculty members are women. Although this is still too low for a multidisciplinary institution, a series of initiatives are gradually redressing the balance.

In Israel, common hurdles for women scientists are magnified because there are only eight major universities and posts are few. These appointments carry an unwritten prerequisite for postdoctoral research experience overseas. But most Israeli PhD graduates are older than their peers abroad — military service is compulsory for women and men, so they often have families and relocation can be a problem.

To encourage more women to stay in academia, the Weizmann Institute launched the Israel National Postdoctoral Program for Advancing Women in Science in 2007. This awards US\$40,000 over two years to supplement the fellowships of ten women graduates selected annually to pursue research overseas. So far, 38 of the 96 recipients have returned to academic positions in Israel; 6 have academic posts abroad; and 5 returned to non-faculty appointments (47 are still overseas).

This year, the Weizmann Institute set up an annual award of \$20,000 over two years for female

postdocs intending to split their research between the institute and a foreign lab. Several Israeli universities and the country's Council for Higher Education now run similar programmes.

Daniella Goldfarb *Weizmann Institute of Science, Rehovot, Israel.*
daniella.goldfarb@weizmann.ac.il

Limit uncertainties in land emissions

Launching satellites to measure carbon dioxide emissions is only part of a more integrated solution to achieving the goals of the Paris climate agreement (*Nature* **533**, 446–447; 2016). In our view, the biggest hurdle is to reduce the high uncertainty in CO₂ emission estimates — particularly from land use.

Land-use emissions are harder to quantify accurately than are those from, say, fossil fuels. To reduce uncertainties in tracking carbon from land use, it is crucial to monitor other data sources such as biomass. This can be done by remote sensing, for example with NASA's GEDI LIDAR sensor or the European Space Agency's proposed BIOMASS (P-band radar) sensor.

We also need many more ground-based measurements of biomass and CO₂ exchange with the atmosphere. However, space agencies' budgets for calibrating and validating satellite products are limited, and there is inadequate coordination and data sharing between the remote-sensing and ground-based measurement communities (see A. K. Skidmore *et al. Nature* **523**, 403–405; 2015).

We suggest that crowdsourced data from mobile-phone apps and more extensive sharing of ground-based measurements could have the greatest potential for improving the monitoring of biomass and CO₂ exchange — at a fraction of the cost of satellites.

Steffen Fritz, Dmitry Schepaschenko, Linda See *IIASA, Austria.*
fritz@iiasa.ac.at

Thomas Kibble

(1932–2016)

Theoretical physicist and Higgs–boson pioneer.

Tom Kibble contributed to our deepest understanding of the fabric and forces of the Universe. He is best known for his work on the phenomenon called spontaneous symmetry breaking — a cornerstone of the standard model of particle physics. His work led to the concept of a new elementary particle now known as the Higgs boson, which was experimentally observed in 2012. Kibble was also a pioneer in applying ideas from both high-energy physics and condensed-matter physics to the study of the early Universe.

Thomas Walter Bannerman Kibble, who died on 2 June, was born in 1932 in Madras (now Chennai), India. His father was a professor of mathematics. He was sent to Edinburgh in 1944 to complete his secondary education at Melville College. Between 1951 and 1958, he pursued a degree in physics and a PhD in mathematical physics at the University of Edinburgh. During the final year of his PhD he married Anne Allan, with whom he had three children. They were happily married until her death in 2005.

In 1959, Kibble joined the theoretical-physics group at Imperial College London, starting an association with the institution that would last for nearly 60 years. It was in the summer of 1964 that, along with US physicists Gerald Guralnik and Carl Richard Hagen, he wrote a seminal paper on symmetry breaking (G. S. Guralnik *et al.* *Phys. Rev. Lett.* **13**, 585–587; 1964).

Symmetry and the breaking of symmetry are deep principles that arise in many different physical contexts. In a lump of iron, for example, the interactions of the atoms respect a perfect symmetry between different directions in space. However, when the iron is cooled to below 770 °C it generates a magnetic field that points in a specific direction, selected spontaneously, and the rotational symmetry is said to be spontaneously broken.

During the early 1960s, initial attempts to incorporate spontaneous symmetry breaking into particle physics showed that certain kinds of massless particle would necessarily be predicted. However, Kibble, Guralnik and Hagen studied spontaneous symmetry breaking in the context of a realization of symmetry called gauge symmetry. In so doing, they reached the striking conclusion that, instead, certain kinds of elementary particles called vector bosons would actually acquire mass.

This beautiful mass-generating



mechanism was independently described slightly earlier in 1964, first in a paper by Belgian physicists Robert Brout and François Englert, and then in a second paper by Peter Higgs, who also explicitly noted that the mechanism gives rise to another massive particle, now known as the Higgs boson.

The significance of these elegant ideas was not immediately recognized. The electromagnetic force is mediated by the photon, a vector boson, but photons are massless. By contrast, the early theory of the weak nuclear force, another of the four fundamental forces of nature, did require massive vector bosons, now called the W and Z bosons. In a key single-author 1967 paper, Kibble showed how the symmetry-breaking mechanism can be generalized, and showed, crucially, how it can leave a vector boson, such as the photon, massless while giving masses to others (T. W. B. Kibble *Phys. Rev.* **155**, 1554; 1967).

These insights became central components in the unification of the electromagnetic and weak nuclear forces into the electroweak theory. (Building on earlier work of Sheldon Glashow, this was formulated in 1967 by Abdus Salam and Steven Weinberg.) In 2010, Kibble shared the American Physical Society's J. J. Sakurai Prize for Theoretical Particle Physics with the five other scientists credited with discovering the spontaneous symmetry-breaking mechanism.

In 2013, Englert and Higgs were awarded the Nobel Prize for Physics. Many people in the theoretical-physics community, including Higgs, hoped that Kibble would be

chosen to share this award. Kibble himself, a man of great modesty, never expressed any disappointment.

In 1976, Kibble laid the foundations for a second key phase of his scientific career. Drawing on ideas from both high-energy physics and condensed-matter physics he realized that phase transitions in the early Universe could leave observable cosmological signatures. He showed that for certain conjectured theories of high-energy physics, the mechanism of spontaneous symmetry breaking predicts novel structures, including one-dimensional concentrations of energy that stretch across the cosmos, which he called cosmic strings.

Kibble analysed how cosmic strings evolve after the phase transition and suggested that they could have a significant impact on the history of the early Universe. Measurements of the cosmic microwave background, the relic radiation from the Big Bang, now indicate that cosmic strings are not the dominant mechanism for creating the large-scale structure of galaxies. Nevertheless, the detection of cosmic strings in future experiments would still provide momentous new insights into the fundamental forces.

Among his many honours, Tom was made a Commander of the British Empire in 1998 and was knighted in 2014. He was especially proud of being the first recipient of the *Nature*/NESTA lifetime achievement award for mentoring in 2005.

Tom was quiet and gently spoken, but he had an influential leadership role in UK academia. As head of the physics department at Imperial College from 1983 to 1991, he skilfully steered it through a difficult period of low funding for science in the United Kingdom.

Tom was also an active campaigner. He joined the British Society for Social Responsibility in Science soon after its formation in 1969, and for three years was chair of the organization's national committee. He was also chair of Scientists Against Nuclear Arms from 1985 to 1991.

Tom changed our understanding of the Universe at the deepest level. He was a man of great integrity and was regarded with much admiration and affection by his colleagues and students. ■

Jerome Gauntlett is head of theoretical physics at Imperial College London.
email: j.gauntlett@imperial.ac.uk

THOMAS ANGUS/IMPERIAL COLLEGE LONDON

MICROBIOLOGY

The dark side of antibiotics

Interactions in the gut between host cells and bacteria can determine a state of health or disease. A study investigates how antibiotic treatment can affect host cells in a way that drives growth of pathogenic bacteria. [SEE LETTER P.697](#)

THIBAUT G. SANA & DENISE M. MONACK

An unwanted side effect of antibiotics can be an increase in pathogenic gut bacteria. A study reported on page 697 by Faber *et al.*¹ now shows that increased growth of the gut pathogen *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) after antibiotic treatment in mice is the result of sugar oxidation that is driven by a host enzyme.

The mammalian gastrointestinal tract is colonized by a dense community of resident microbes known as the gut microbiota that not only helps us to digest certain foods, but also helps to prevent colonization by invading and potentially hostile microorganisms — a property known as colonization resistance. Perturbations to the microbiota, such as those caused by the use of oral antibiotics, often lead to increased colonization by various gut pathogens, such as *S. Typhimurium* and *Clostridium difficile*^{2,3}. Broad-spectrum antibiotics deplete the resident (commensal) microbiota, allowing pathogens to proliferate, which in turn can lead to gastrointestinal inflammation⁴. Antibiotic-associated diarrhoea and inflammation occurs in 5–25% of people treated with antibiotics and is considered a major health problem⁵.

Nutrient availability is a key factor in determining bacterial growth. It has been known for decades that pretreatment of mice with the antibiotic streptomycin increases the severity of colon inflammation induced by *S. Typhimurium*⁶, and in the inflamed gut there are multiple nutrients that can facilitate the replication of this bacterium. For example, *S. Typhimurium* and other members of the Enterobacteriaceae family can use ethanolamine derived from the degradation of cell membranes as a source of carbon and nitrogen^{7–9}.

In addition, antibiotic-mediated disruption of the microbial food web can give rise to microbiota-liberated sugars in the gut that promote the growth of *S. Typhimurium* and *C. difficile*¹⁰. Moreover, antibiotics also increase expression of a host enzyme called inducible nitric oxide synthase (iNOS)¹¹, but the link between iNOS and enhanced *S. Typhimurium* growth in the gut had not been previously established. The key advance in the paper by

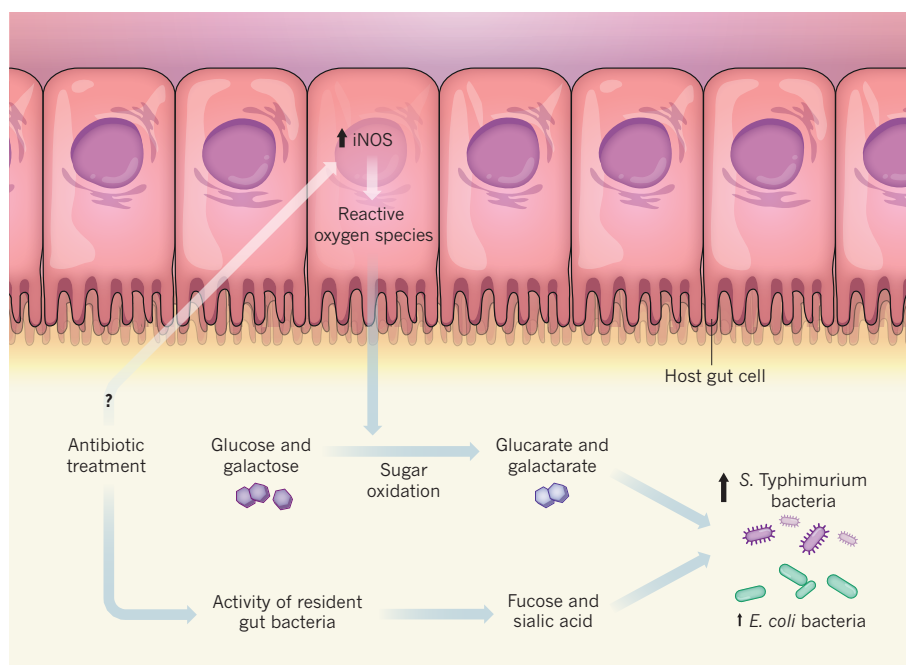


Figure 1 | Gut changes after antibiotic treatment. Faber *et al.*¹ confirmed that treatment of mice with antibiotics increases the expression of the host enzyme inducible nitric oxide synthase (iNOS) by an unknown mechanism. iNOS can produce reactive oxygen species, and the authors propose that such species can oxidize glucose and galactose sugars to glucarate and galactarate, respectively. These oxidized sugars are metabolized by *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*), as well as by other members of the Enterobacteriaceae family, such as resident *Escherichia coli*. Antibiotic treatment also results in the formation of sialic acid and fucose by the activity of resident bacteria; these nutrients can also be metabolized by *S. Typhimurium*¹⁰. Antibiotic treatment therefore ultimately creates a perfect mix of nutrients for *S. Typhimurium* to proliferate.

Faber and colleagues is to provide mechanistic insight into how antibiotic treatment leads to iNOS-dependent generation of oxidized sugars, which *S. Typhimurium* uses as a food source to grow rapidly in the gut.

The authors investigated the changes that occur after antibiotic treatment. After confirming that the treatment leads to overexpression of iNOS (by an unknown mechanism), the authors showed that host-driven iNOS-dependent oxidation of the sugars glucose and galactose in the gut leads to the generation of the sugars glucarate and galactarate, both of which can be metabolized by *S. Typhimurium*.

The researchers then characterized *S. Typhimurium* genes involved in metabolizing oxidized glucose and galactose (the *gudT*

ygY gudD STM2959 operon). The expression of these genes is induced by hydrogen, a fermentation product of the microbiota¹², indicating that the production of the enzymes that they encode is tightly regulated, and probably restricted to a gut-like environment. Interestingly, related genes were found in other Enterobacteriaceae present in the resident bacterial community, such as *Escherichia coli* and *Klebsiella oxytoca*. The authors show that these genes are also important for *E. coli* to increase in numbers in the gut after antibiotic treatment.

Because *S. Typhimurium* probably competes with commensal bacteria for the oxidized sugars, it is tempting to speculate that it might have evolved specific mechanisms to outcompete resident bacteria for the same food source. The

elucidation of such competition mechanisms would add a new layer of complexity to our understanding of the microbiota's response to antibiotics and will require further studies. Thus, the increase of *S. Typhimurium* in the gut after antibiotic treatment can be attributed to the microbiota-delivered nutrients sialic acid and fucose¹⁰ and to host-mediated oxidation of carbohydrates in the gut, providing diverse food sources for the pathogen (Fig. 1).

Antibiotics are useful for treating susceptible bacterial infections and certainly provide human health benefits. However, with the emergence of multidrug-resistant pathogens that are predicted¹³ to kill 10 million people a year by 2050, there is a dark side to antibiotics. The mechanism described by Faber *et al.* sheds light on another dark side. Indeed, increasing

numbers of studies now describe mechanisms used by gut pathogens to take advantage of the post-antibiotic period to hyper-replicate within the gut and successfully infect the host. Faber and colleagues' work could lead to the development of new and better therapeutic approaches for preventing diseases that are a consequence of antibiotic treatment. ■

Thibault G. Sana and Denise M. Monack
are in the Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.
e-mails: tsana@stanford.edu;
dmonack@stanford.edu

1. Faber, F. *et al.* *Nature* **534**, 697–699 (2016).
2. Pavia, A. T. *et al.* *J. Infect. Dis.* **161**, 255–260 (1990).

3. Kelly, C. P., Pothoulakis, C. & LaMont, J. T. *N. Engl. J. Med.* **330**, 257–262 (1994).
4. Högenauer, C., Hammer, H. F., Krejs, G. J. & Reisinger, E. C. *Clin. Infect. Dis.* **27**, 702–710 (1998).
5. Bergogne-Bérézin, E. *Int. J. Antimicrob. Agents* **16**, 521–526 (2000).
6. Bohnhoff, M., Drake, B. L. & Miller, C. P. *Proc. Soc. Exp. Biol. Med.* **86**, 132–137 (1954).
7. Kofoed, E., Rappleye, C., Stojiljkovic, I. & Roth, J. *J. Bacteriol.* **181**, 5317–5329 (1999).
8. Garsin, D. A. *Nature Rev. Microbiol.* **8**, 290–295 (2010).
9. Thiennimitr, P. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 17480–17485 (2011).
10. Ng, K. M. *et al.* *Nature* **502**, 96–99 (2013).
11. Songhet, P. *et al.* *PLoS ONE* **5**, e13804 (2010).
12. Lamichhane-Khadka, R., Benoit, S. L., Maier, S. E. & Maier, R. J. *Open Biol.* **3**, 130146 (2013).
13. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations* (Wellcome Trust/UK Gov., 2016).

This article was published online on 15 June 2016.

PHYSIOLOGY

Stressed-out chromatin promotes longevity

Two studies reveal that early-life malfunction in organelles called mitochondria brings about lasting changes in how DNA is packaged. These alterations have consequences for cellular stress responses and organismal longevity.

SIU SYLVIA LEE & JESSICA K. TYLER

Organelles called mitochondria are central to the generation of ATP molecules, which are the energy currency of cells. Mild mitochondrial impairment in juvenile roundworms (*Caenorhabditis elegans*) has been shown to increase lifespan^{1,2}. Such dysfunction triggers multiple signalling cascades, including a pathway called the mitochondrial unfolded protein response³ (UPR^{mt}), which causes changes in gene expression that enable cells and organisms to cope with stress⁴. Writing in *Cell*, Tian *et al.*⁵ and Merkwirth *et al.*⁶ identify factors that relay mitochondrial dysfunction to the UPR^{mt}.

Changes in gene expression are largely brought about by alterations in how tightly DNA is packaged around histone proteins in a structure called chromatin. Chromatin regulation is achieved partly through chemical decorations that modify histones, modulating this packaging. The current studies reveal that histone-modifying enzymes are integral players in the UPR^{mt} stress response and lifespan extension in *C. elegans*.

By using genetic screens, Tian and colleagues found that the gene *lin-65* is essential for the induction of UPR^{mt} in response to mitochondrial stress. However, it should be noted that lifespan was extended even in the absence of *lin-65*, suggesting that mitochondrial stress

also induces other pathways that promote longevity. The authors demonstrated that mitochondrial stress causes the LIN-65 protein to shuttle from the cytoplasm to the nucleus, reminiscent of the UPR^{mt} regulator and transcription factor DVE-1, which also undergoes stress-induced nuclear migration⁷. Indeed, the researchers found that LIN-65 is required for the nuclear relocalization of DVE-1.

Tian *et al.* provided evidence to suggest that the addition of two methyl groups to amino-acid residue lysine 9 of histone H3 (a modification dubbed H3K9me2) in the cytoplasm by the MET2 methyltransferase enzyme triggered DVE-1 to shuttle to the nucleus. However, the possibility that non-histone targets of methylation are involved in this process cannot be ruled out, because the authors' suggestion is based on methyltransferase inactivation, which could alter the methylation of many other proteins besides histones. The mechanism by which H3K9me2 promotes DVE-1 nuclear migration remains unknown.

The H3K9me2 modification is generally associated with tightly packaged chromatin. Indeed, Tian and colleagues observed that nuclei in worms exposed to mitochondrial stress consistently looked smaller and more dense than those in unstressed worms. DVE-1 localized to discrete spots away from DNA-dense regions, suggesting that this factor preferentially targets DNA regions that

are less tightly packaged. This work supports a model whereby stress, through H3K9me2, induces changes in chromatin packaging that restrict the number of DNA regions to which DVE-1 can bind to regulate transcription. Thus, stress leads to specific changes in gene expression (Fig. 1).

In a complementary study, Merkwirth *et al.* identified the genes *jmjd-1.2* and *jmjd-3.1* as crucial for the activation of the UPR^{mt} pathway and extended lifespan in worms that had mitochondrial impairment. These genes encode enzymes that remove methyl groups from a histone modification dubbed H3K27me3, which is associated with tightly packaged chromatin. Mitochondrial impairment increases the expression of these two histone demethylases, which presumably leads to unpackaging of chromatin (Fig. 1). Overexpression of *jmjd-1.2* and *jmjd-3.1* in *C. elegans* recapitulates many of the effects of mild mitochondrial stress, including increased longevity and UPR^{mt} induction, perhaps because increased levels of JMJD-1.2 and JMJD-3.1 repress the expression of many mitochondrial genes, thereby impairing mitochondrial function.

In both studies, the histone modifiers seemed to exert their effects predominantly on juvenile worms. It is therefore tempting to speculate that chromatin is most receptive to stress-induced reorganization in early life, and that these changes help to establish a stress response that persists into adulthood. Histone modifications can be stably maintained through cell divisions, even through generations, making them good candidate substrates for a 'cellular memory' of mitochondrial dysfunction. It will be interesting to determine whether H3K9me2 methyltransferases and H3K27me3 demethylases are required for maintenance of the mitochondrial-dysfunction-induced transcriptional program and activation of the UPR^{mt} through to adulthood, even when the mitochondrial stress that triggered it is removed.

Because regulation of gene expression is

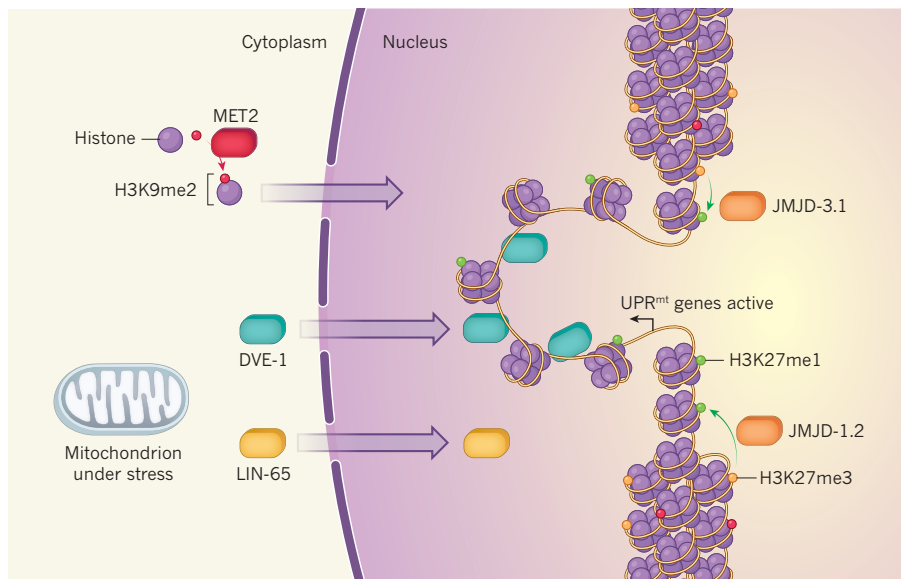


Figure 1 | Coordinated responses to mitochondrial stress. Gene expression can be inhibited by the tight packaging of DNA as chromatin around histone proteins that are decorated with methyl groups at specific amino-acid residues. Two such repressive modifications are dubbed H3K9me2 and H3K27me3. Two studies^{5,6} reveal that mitochondrial stress triggers the unpacking and expression of genes involved in a stress response called the UPR^{mt}. Tian *et al.*⁵ find that movement of the protein LIN-65 from the cytoplasm to the nucleus triggers nuclear migration of the transcription factor DVE-1. The enzyme MET2 mediates the addition of H3K9me2 modifications to histones in the cytoplasm; these histones move to the nucleus, restricting the unpacked regions where DVE-1 can bind to promote transcription. Merkwirth *et al.*⁶ show that JMJD-1.2 and JMJD-3.1 enzymes are upregulated in response to mitochondrial stress and remove methyl groups from H3K27me3 to form the transcription-promoting modification H3K27me1 in the vicinity of UPR^{mt} genes.

already known to be central to the response to mitochondrial stress⁸, the participation of histone modifiers is not in itself unexpected. However, the current studies suggest that chromatin-modifying enzymes can be highly selective, eliciting programs of gene expression that are specific to a particular stress or pathway. Patterns of histone modifications can be extremely diverse, so it is not difficult to imagine how such exquisite specificity can arise. There are hundreds of histone modifications and many enzymes that have distinct and overlapping roles in depositing and removing them. DNA is wound around a nucleosome complex composed of four different histones, and each histone can be modified at many positions, so the number of possible combinations of histone modifications on each nucleosome is huge. Moreover, each nucleosome can have distinct modifications from its neighbours, building up an astronomically complex 'histone code'.

Both groups found that the regulation of chromatin in response to mitochondrial stress is tissue specific. Tian *et al.* showed that LIN-65 activates the UPR^{mt} pathway in the intestine, whereas Merkwirth *et al.* reported that JMJD-1.2 and JMJD-3.1 trigger the same response in neurons. An intriguing possibility is that distinct chromatin-regulating factors in different cells and tissues sense mitochondrial stress and respond in divergent yet coordinated ways that best enable organisms to counter the stress. It will be crucial to elaborate

how stress signals can be coupled to chromatin regulation, and how cell-type-specific regulation can be coordinated. Perhaps mitochondrial impairment activates signalling cascades that regulate the expression, movements, modifications and activities of distinct histone modifiers in specific cells. Mitochondrial impairment could also result in altered levels of metabolic

intermediates, many of which act as essential co-factors of histone-modifying enzymes.

The idea that a stress signal can engage specific histone modifiers to elicit a persistent response will no doubt extend to diverse stressors. Indeed, JMJD-3.1 has been implicated⁹ in the dampening of a stress response called the heat-shock response from early adulthood through to ageing in *C. elegans*. Additionally, by-products of cellular metabolism called free radicals modulate longevity in yeast through a histone-modifying enzyme¹⁰. We can look forward to many more exciting insights into the role of chromatin regulation in stress responses. ■

Siu Sylvia Lee is in the Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14850, USA.

Jessica K. Tyler is in the Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York 10065, USA.

e-mails: sylvia.lee@cornell.edu;
jet2021@med.cornell.edu

1. Dillin, A., Crawford, D. K. & Kenyon, C. *Science* **298**, 830–834 (2002).
2. Rea, S. L., Ventura, N. & Johnson, T. E. *PLoS Biol.* **5**, e259 (2007).
3. Haynes, C. M. & Ron, D. J. *Cell Sci.* **123**, 3849–3855 (2010).
4. Durieux, J., Wolff, S. & Dillin, A. *Cell* **144**, 79–91 (2011).
5. Tian, Y. *et al.* *Cell* **165**, 1197–1208 (2016).
6. Merkwirth, C. *et al.* *Cell* **165**, 1209–1223 (2016).
7. Haynes, C. M., Petrova, K., Benedetti, C., Yang, Y. & Ron, D. *Dev. Cell* **13**, 467–480 (2007).
8. Chang, H.-W., Shtessel, L. & Lee, S. S. *Free Radic. Biol. Med.* **78**, 168–178 (2015).
9. Labbadia, J. & Morimoto, R. I. *Mol. Cell* **59**, 639–650 (2015).
10. Schroeder, E. A., Raimundo, N. & Shadel, G. S. *Cell Metab.* **17**, 954–964 (2013).

EARTH SCIENCE

An extended yardstick for climate variability

Decoded and precisely dated information encrypted in stalagmites from a cave in China reveal past climatic changes and provide insight into the complex interactions in today's climate system. SEE LETTER P.640

NELE MECKLER

Cave stalagmites represent a meticulous archive of climatic events stretching back over many hundreds of thousands of years and, unlike many geological deposits on land and in the ocean, allow the timing of these events to be precisely determined. On page 640 of this issue, Cheng *et al.*¹ decipher a treasure trove of information concealed

in the stalagmites of the Sanbao Cave in the mountains of central China. Their findings extend one of the most valuable benchmark climate records available, which now spans an impressive 640,000 continuous years of Asian monsoon history with unprecedented dating precision. The Asian monsoon is a crucial component of the global climate system, so these data not only provide age constraints for other palaeoclimate time series, but also



Figure 1 | Stalagmites in Reed Flute Cave, Guilin, China. Mineral deposits such as stalagmites can be used as a record of climate.

improve our understanding of the interactions, forcing factors and responses of today's climate systems.

As they grow upwards from the cave floor, stalagmites accumulate layer upon layer of calcium carbonate (Fig. 1). Deposited from oversaturated drip waters, this compound is imprinted with the water's chemical composition, which, in turn, carries the chemical signature of rainwater falling above the cave. The oxygen isotope composition of this water (heavy ^{18}O compared with light ^{16}O , reported as $\delta^{18}\text{O}$) is an indicator of the climate conditions at the time the rain fell (Fig. 2). Dissolved uranium leached from the rocks above the cave is also trapped in the forming calcium carbonate, acting as a radiometric clock as it slowly decays. This allows stalagmite age to be accurately determined back to around 650,000 years ago². The older the material, the less exact this clock becomes, and the more effort is needed to determine precise age constraints. Cheng *et al.* go to the limit of what can be achieved, and their precisely dated stalagmite record covers more than six glacial-to-interglacial cycles.

Climatic events are often globally interconnected, sometimes to a surprising extent. For example, during glacial periods there is a remarkable synchrony between dry events in the Asian monsoon climate and cold events in the North Atlantic region: this is related to massive discharge of icebergs and changes in ocean and atmospheric circulation^{3,4}. The repercussions extend as far south as Antarctica, which warms during these events^{5,6}, and include alterations in the concentrations of methane and carbon dioxide

in the atmosphere. Such connections are well established for periods during which good age control is available for all archives, and can provide crucial lessons about our climate system. Under the (reasonably safe) assumption that the synchrony observed for the more recent past holds further back in history, the extended cave record described by Cheng *et al.* can be used as an indirect age constraint for other archives^{3,7} that lack direct radiometric age control. This information therefore greatly extends the time over which we have precisely dated climate data around the globe.

It might be asked why long time series are important, given that it is so much easier to obtain well-dated climate records for the more recent past. The issue is the multitude of contributing factors that coincide at any one time. For example, Earth's climate is strongly influenced by the configuration of its orbit relative to the Sun, which changes cyclically. The time of the year at which Earth is closest to the Sun changes every 20,000 years or so (the precession cycle). Cyclical changes in the tilt of Earth's axis (the obliquity cycle) occur with a periodicity of around 40,000 years, leading to additional variations in the strength of winter and summer seasons at high latitudes⁸. Which of these two cyclical variations is responsible for the observed duration of glacial cycles of roughly 100,000 years has long been a matter of debate^{9–11}.

The answer to this question can come only from a well-dated climate time series that covers a sufficient number of these cycles — like that amassed by Cheng and colleagues. Although this record reflects monsoon strength and not the extent of glaciation,



50 Years Ago

Atlantic tunas will share with whales some immunity from hunting as a result of a convention on conservation drafted ... by the representatives of seventeen nations ... The convention has been prompted by evidence in recent years that more boats are chasing fewer fish. It applies to the whole of the Atlantic Ocean and adjacent seas, and the hope is that it will maintain the population of tuna ... at those levels which permit the greatest sustainable catch. The commission that will be responsible for administering the convention ... will be empowered to collect information and to recommend research programmes. It may also, on the basis of the evidence it collects, recommend limits on the rate of catch in the convention area.

From Nature 25 June 1966

100 Years Ago

I refer to the continued proposals to found fresh scholarships for the encouragement of scientific research, accompanied as they so often are by statements as to the lack of trained men in science ... The men who have held ... scholarships for two or three years form a body highly trained in the best English and Continental universities ... Yet we see on all hands these men barely able to make a living ... They are in general men of all-round education, with specialised knowledge in science in addition; they are not particularly uncouth, unpractical, or unbalanced, as popular tradition would have men of science to be. It is this addition of specialised knowledge that ... is the greatest obstacle to their earning a living; they would probably be better paid if they turned their hand to any employment other than the pursuit of science. **E. N. da C. Andrade, British Expeditionary Forces**

From Nature 29 June 1916

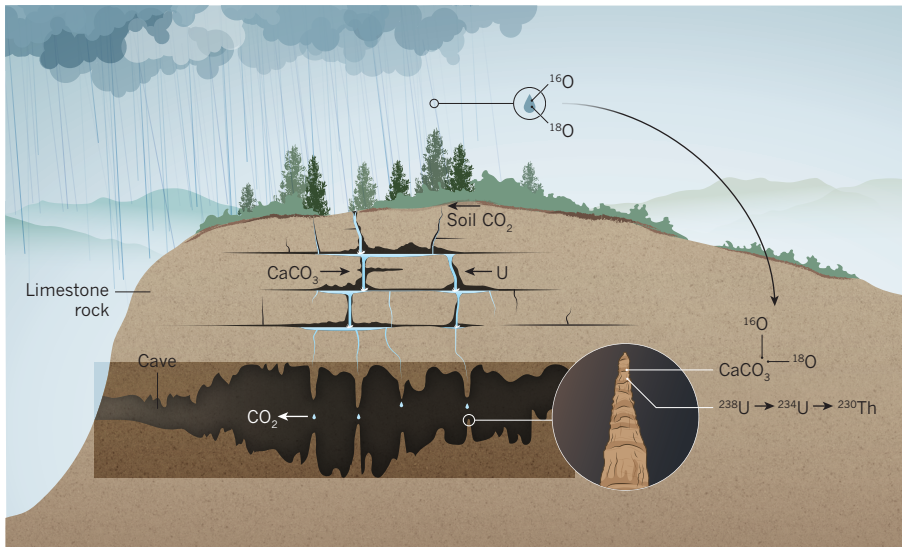


Figure 2 | Archiving of climate events. Cheng *et al.*¹ measure the relative proportions of the oxygen isotopes ^{16}O and ^{18}O in stalagmites to determine the intensity of past monsoons in central China. Rainwater acidified by soil carbon dioxide leaches calcium carbonate (CaCO_3) and traces of radioactive uranium (U) from limestone rock above the cave. As it percolates into the cave, the water loses part of its CO_2 , causing calcium carbonate to precipitate out, forming stalactites and stalagmites on the roof and floor of the cave, respectively. This carbonate retains the rainwater's original oxygen-isotope composition — an indicator of the climate conditions when the rain fell¹² — and the dripwater's uranium. As the uranium slowly decays to thorium (Th), it acts as a record of the time since calcite deposition. Cheng *et al.* use these processes to compile a remarkable time series of monsoon variability over 640,000 years.

a previously established link between particularly strong dry events and glacial terminations allowed the authors to determine the exact timing of the past seven glacial terminations. These all relate to specific times in the precession cycle and occur every four or five precession cycles, revealing a prominent role for this orbital cycle.

In general, the extended cave record shows

that the pattern of monsoon variability seen in the more recent past also occurred in earlier periods, indicating that it is an inherent feature of natural climate variability. Cyclical variation in monsoon strength in relation to orbital cycles is particularly prominent. Furthermore, dry events of shorter duration are evident throughout the record. Cheng *et al.* investigate these dry periods by removing the orbitally

controlled cyclicality from their monsoon record and studying the remaining suborbital signals (termed $\Delta\delta^{18}\text{O}$). This approach could create artefacts, a possibility investigated by Cheng *et al.* (see Extended Data, Figs 5, 6), but allows the suborbital features to stand out more clearly in this residual time series — and again, these seem to be related to orbital variations in the precession cycle. If it holds up, this could be an invaluable clue about the unknown cause of such short-lived events.

The extended time series that Cheng *et al.* derive from the Sanbao Cave stalagmites has an array of other interesting features that call for further investigation. This wealth of new information on past climate variability over an exceptionally long time period and with precise age control will bring us a step closer to understanding the drivers and interactions in our climate system. ■

Nele Meckler is in the Department of Earth Science and Bjerknes Centre for Climate Research, University of Bergen, 5007 Bergen, Norway.

e-mail: nele.meckler@uib.no

1. Cheng, H. *et al.* *Nature* **534**, 640–646 (2016).
2. Cheng, H. *et al.* *Earth Planet. Sci. Lett.* **371–372**, 82–91 (2013).
3. Cheng, H. *et al.* *Science* **326**, 248–252 (2009).
4. Wang, Y. J. *et al.* *Science* **294**, 2345–2348 (2001).
5. Stocker, T. F. *Science* **282**, 61–62 (1998).
6. EPICA Community Members *Nature* **444**, 195–198 (2006).
7. Barker, S. *et al.* *Science* **334**, 347–351 (2011).
8. Ruddiman, W. F. *Earth's Climate: Past and Future* 3rd edn (Freeman, 2013).
9. Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
10. Raymo, M. E. *Paleoceanography* **12**, 577–585 (1997).
11. Huybers, P. *Nature* **480**, 229–232 (2011).
12. Dansgaard, W. *Tellus* **16**, 436–468 (1964).

host's body. Until now, such transmissible cancers — cancer-cell lineages with the potential to metastasize through an animal population — were considered to be exceedingly rare. Only four examples were known in nature: two affecting Tasmanian devils, one in dogs and another in soft-shell clams^{2–4}. Metzger and co-workers now report four previously unidentified transmissible cancers: one that affects mussels (*Mytilus trossulus*) found in British Columbia, one that affects golden carpet shell clams (*Polititapes aureus*) on the Iberian coast and two transmissible cancers of probably independent origin in common cockles (*Cerastoderma edule*).

These cancers all cause a leukaemia-like disease in affected individuals called disseminated neoplasia, which had previously been observed, and which manifests as an excess of large, abnormal cells in the circulatory system. Diseased animals have thick, opaque circulatory fluid, and their tissues become clogged with invasive cancer cells^{5,6}. The tendency for

CANCER

Transmissible tumours under the sea

In some species, cancer cells can be directly transmitted between individuals. An analysis in shellfish now shows that some transmissible cancers can even cross the species barrier. SEE LETTER p.705

ELIZABETH P. MURCHISON

On page 705, Metzger *et al.*¹ report the discovery that transmissible cancers are widespread in one group of marine shellfish, known as bivalves, and that such cancers can even jump between species. These findings suggest that cancer cells are common infectious agents in marine environments, and

challenge our understanding of the nature of cancer and its interaction with its hosts.

Cancer occurs when a single cell in the body acquires genetic changes that drive inappropriate cell proliferation. Once initiated, cancer evolves by natural selection, often producing cell lineages that spread through the host by a process called metastasis. However, cancer does not normally spread beyond the

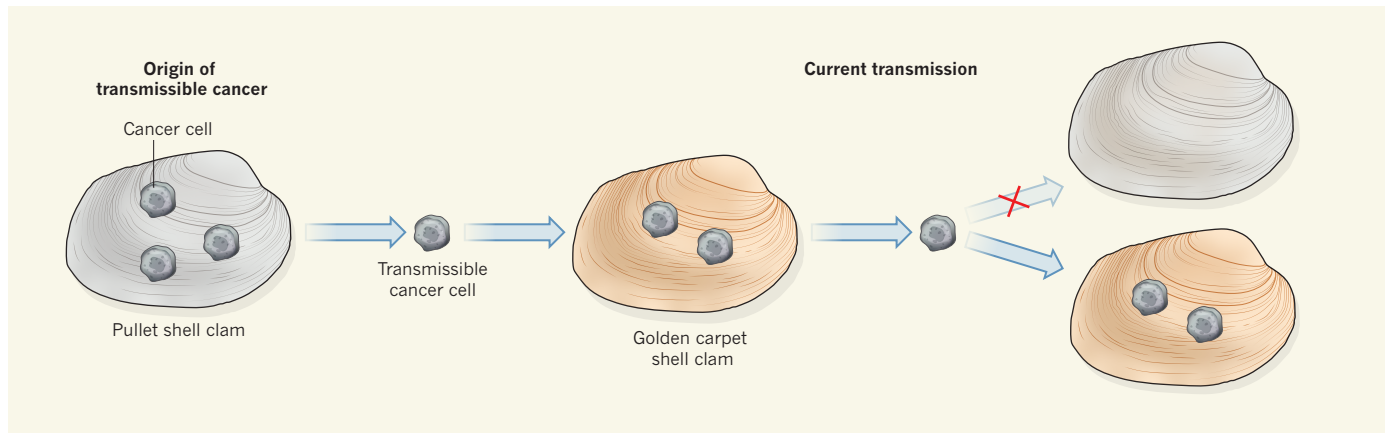


Figure 1 | Cancer cells can be transmitted between shellfish species. A study by Metzger *et al.*¹ has found that a transmissible cancer occurring in golden carpet shell clams (*Polittapes aureus*) originated in a different species — the pullet shell clam (*Venerupis corrugata*). Although the two clam species share a habitat, the cancer is currently detected only in golden carpet shell clams, suggesting that pullet shell clams may have acquired resistance to infection by this cancer.

many bivalve species to develop disseminated neoplasia has been known⁷ since the 1960s, but the underlying cause of the condition was not understood.

Metzger and colleagues performed a genetic analysis of cancer and host tissues from several individual mussels, cockles and golden carpet shell clams. They found that, in many cases, the cancer cells bore no genetic similarity to their hosts, but instead were highly similar to cancerous tissues derived from other individuals of the same bivalve species. These findings confirmed that many cases of disseminated neoplasia in bivalve species are due to horizontal transfer of living cancer cells between hosts.

A particularly unexpected finding of Metzger and colleagues' work was that DNA extracted from cancer cells in golden carpet shell clams showed no genetic match with normal DNA from this species, but instead indicated that the cancer cells originated in a different species — the pullet shell clam (*Venerupis corrugata*). Surprisingly, however, pullet shell clams — which share a habitat with golden carpet shell clams — are not known to have a high prevalence of disseminated neoplasia. Perhaps the pullet shell clam has adapted to resist infection by the transmissible cancer that first arose in a member of its own species; despite this, the cancer has survived by engrafting to a new host species (Fig. 1).

Altogether, these findings seem to paint a picture of shellfish beds around the world that are awash with microscopic cancer cells metastasizing both within and between species. Although the mechanisms of cancer transmission remain unclear, the immobile nature of these filter-feeding invertebrates suggests that the cancer cells may float through the marine environment and enter their hosts by breaching the digestive or respiratory tracts. The mode by which cancer cells exit their diseased hosts is another puzzle. Perhaps this is a passive process enabled by trauma or predation,

or maybe cancer cells actively migrate out of the body by co-opting host signalling pathways. Investigating the density and viability of free-living bivalve neoplastic — and non-neoplastic — cells in the external marine environment will be an interesting area for future study.

Although disseminated neoplasia has been reported in many bivalve species, the current work and previous studies^{5,6} reveal that its prevalence varies greatly both within and between species. Variable prevalence of bivalve transmissible cancers, particularly within localized populations, hints at a fierce, ongoing pathogen–host evolutionary arms race beneath the sea. Although any mechanisms of host immunity to the cancers are unknown, their elucidation will provide insight into the diversity of cancer immunological and immune-evasion processes across species. Furthermore, it is not known how frequently new disseminated neoplasias arise in bivalves; identifying the genetic changes that distinguish cancers that remain in one host from those that become transmissible may provide valuable information about the mechanisms of transmissibility.

Determining the timescales and geographical distances that underpin the evolutionary histories of transmissible cancers in bivalves will provide a greater understanding of these diseases. It is possible that, like the canine transmissible cancer⁸, these cancers are ancient cell lineages that have co-evolved with their hosts through the millennia; or perhaps their emergence is a relatively recent occurrence, possibly stimulated by infectious agents, environmental changes, aquaculture or other anthropogenic activities.

The potential for cancer cells to become free-living infectious agents raises questions about the implications for cancer transmission in humans. Although person-to-person transmission and survival of cancer cells has been reported during organ transplantation,

pregnancy, experimental treatments and surgical accidents, such exchanges are rare and never spread beyond transfer between two individuals⁸. Interestingly, however, the recent discovery of tapeworm neoplastic cells that spread within their severely immunocompromised human host⁹ consolidates Metzger and co-workers' finding that cancers can invade new host species.

The risk of cancer is inherent in multicellular organisms, and the basic evolutionary drive of this disease does not respect individual or even species barriers. Bivalve transmissible cancers provide a new model system in which to explore cancer transmission and host response. An understanding of the aetiology of disseminated neoplasias in these animals is also a boon for the aquaculture industry, providing new opportunities for disease biomonitoring and control. The discovery of widespread transmissible cancers under the sea is an exciting conceptual advance, and opens up further avenues for cancer research. ■

Elizabeth P. Murchison is in the Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, UK. e-mail: epm27@cam.ac.uk

1. Metzger, M. J. *et al.* **534**, 705–709 (2016).
2. Metzger, M. J., Reinisch, C., Sherry, J. & Goff, S. P. *Cell* **161**, 255–263 (2015).
3. Murchison, E. P. *Oncogene* **27** (Suppl. 2), S19–S30 (2008).
4. Pye, R. J. *et al.* *Proc. Natl Acad. Sci. USA* **113**, 374–379 (2016).
5. Barber, B. J. *Aquat. Living Resour.* **17**, 449–466 (2004).
6. Carballal, M. J., Barber, B. J., Iglesias, D. & Villalba, A. *J. Invert. Pathol.* **131**, 83–106 (2015).
7. Fanley, C. A. *Natl Cancer Inst. Monogr.* **31**, 541–555 (1969).
8. Strakova, A. & Murchison, E. P. *Curr. Opin. Genet. Dev.* **30**, 49–55 (2015).
9. Muehlenbachs, A., Mathison, B. A. & Olson, P. D. *N. Engl. J. Med.* **373**, 1845–1852 (2015).

This article was published online on 22 June 2016.

Paris Agreement climate proposals need a boost to keep warming well below 2 °C

Joeri Rogelj^{1,2}, Michel den Elzen³, Niklas Höhne^{4,5}, Taryn Fransen⁶, Hanna Fekete⁴, Harald Winkler⁷, Roberto Schaeffer⁸, Fu Sha⁹, Keywan Riahi^{1,10} & Malte Meinshausen^{11,12}

The Paris climate agreement aims at holding global warming to well below 2 degrees Celsius and to “pursue efforts” to limit it to 1.5 degrees Celsius. To accomplish this, countries have submitted Intended Nationally Determined Contributions (INDCs) outlining their post-2020 climate action. Here we assess the effect of current IND Cs on reducing aggregate greenhouse gas emissions, its implications for achieving the temperature objective of the Paris climate agreement, and potential options for overachievement. The IND Cs collectively lower greenhouse gas emissions compared to where current policies stand, but still imply a median warming of 2.6–3.1 degrees Celsius by 2100. More can be achieved, because the agreement stipulates that targets for reducing greenhouse gas emissions are strengthened over time, both in ambition and scope. Substantial enhancement or over-delivery on current IND Cs by additional national, sub-national and non-state actions is required to maintain a reasonable chance of meeting the target of keeping warming well below 2 degrees Celsius.

In December 2015, the Paris Agreement¹—a new global agreement to combat climate change—was adopted under the United Nations Framework Convention on Climate Change (UNFCCC). In preparation of this agreement, countries submitted national plans that spell out their intentions for addressing the climate change challenge after 2020². These Intended Nationally Determined Contributions (INDCs) address a range of issues, which can relate to avoiding, adapting or coping with climate change, among other things. Nevertheless, targets and actions for reducing greenhouse gas (GHG) emissions are core components. At this point, the IND Cs are not final and can be modified up until the time the Paris Agreement is ratified. However, for now they represent our best understanding of the climate actions countries intend to pursue after 2020.

The overarching climate goal of the Paris Agreement is to hold “the increase in the global average temperature to well below 2 °C above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 °C above pre-industrial levels”¹. This climate goal represents the level of climate change that governments agree would prevent dangerous interference with the climate system, while ensuring sustainable food production and economic development^{3,4}, and is the result of international discussions over multiple decades⁵. Limiting warming to any level implies that the total amount of carbon dioxide (CO₂) that can ever be emitted into the atmosphere is finite⁶. From a geophysical perspective, global CO₂ emissions thus need to become net zero^{7,8}. About two thirds of the available budget for keeping warming to below 2 °C have already been emitted^{9–11}, and increasing trends in CO₂ emissions¹² indicate that global emissions urgently need to start to decline so as to not foreclose the possibility of holding warming to well below 2 °C (refs 13, 14). The window for limiting warming to below 1.5 °C with high probability and without temporarily exceeding that level already seems to have closed¹⁵. The Paris Agreement implicitly acknowledges these insights and aims to reach a global peak in GHG emissions as soon as possible together with achieving “a balance” between anthropogenic emissions and removals of GHGs in the second half of this century. Both targets are in principle

consistent with the temperature objective of the Agreement^{16,17}, but beg the broader question of whether current IND Cs are already putting the world on a path towards achieving them.

Besides the climate question, the first round of IND Cs also raises many other issues. These include whether efforts are distributed equitably among countries; how much adaptation may be required given the current level of mitigation ambition; how ‘intended’ national proposals will be implemented; how they will be financed; and the extent to which the IND Cs contribute to the achievement of other goals of the UNFCCC by building on institutions that can support adaptation to climate change, technology advancement, development path transformation, sustainable development and enhanced awareness. Although these issues are important for many countries, they fall outside the scope of this analysis.

In this Perspective, we assess the implications of the current IND Cs for GHG emissions, including the main factors and uncertainties that influence the levels of GHG emissions in 2030—the latest year covered by the vast majority of IND Cs—and we explore the consistency of these reductions with the objective of the Paris Agreement (to keep warming well below 2 °C and pursue efforts towards 1.5 °C). This work updates and expands work undertaken in the framework of the 2015 United Nations Environment Programme (UNEP) Emissions Gap Report¹⁸—an authoritative annual assessment that has tracked climate policy action over the past six years, and provides a synthesis of a wide range of INDC modelling studies^{19–29} that are available in the public domain. The number of IND Cs considered by the studies that we assess here ranges from the 118 IND Cs submitted by 1 October 2015 to the 160 IND Cs submitted by 12 December 2015 (Supplementary Tables 1 and 2). These 118 to 160 IND Cs cover emissions from 145 to 187 out of 195 Parties to the UNFCCC, which in turn were responsible for roughly 88% to more than 96% of global GHG emissions in 2012³⁰. We also look at projections of global-mean temperature increase over the twenty-first century that would be consistent with the IND Cs, and at post-2030 implications for limiting warming to no more than 2 °C. Finally, we discuss options to

¹ENE Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. ²Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. ³PBL Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands. ⁴NewClimate Institute, Cologne, Germany. ⁵Environmental Systems Analysis Group, Wageningen University, Wageningen, The Netherlands. ⁶World Resources Institute, Washington DC, USA. ⁷Energy Research Center, University of Cape Town, Cape Town, South Africa. ⁸Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brazil. ⁹National Center for Climate Change Strategy and International Cooperation, Beijing, China. ¹⁰Graz University of Technology, Graz, Austria. ¹¹Australian-German Climate and Energy College, School of Earth Sciences, The University of Melbourne, Melbourne, Victoria, Australia. ¹²PRIMAP Group, Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany.

BOX 1

Scenario definitions

Scenarios represent alternative images of the future, or “[stories] about what happened in the future”⁷⁶. They are neither predictions nor forecasts⁷⁷, but tools to understand how the future might unfold under a consistent set of assumptions. In this analysis, we use four types of scenarios, drawn from a wide variety of sources.

No-policy baseline scenarios. These are emissions projections that assume that no new climate policies have been put into place from 2005 onwards. We select these scenarios from the IPCC AR5 Scenario Database¹⁶, which is hosted at the International Institute for Applied Systems Analysis (IIASA, <https://tntcat.iiasa.ac.at/AR5DB/>). By design, these no-policy baseline scenarios exclude climate policies, but may include other policies that can influence emissions and are implemented for other reasons, such as energy efficiency or energy security policies.

Current-policy scenarios. These consider the most recent estimates of global emissions and take into account implemented national policies. This is different from the INDC scenarios (described below), which reflect international pledges and intended policies. Here, we draw these scenarios from three global analyses^{19,29,78}.

INDC scenarios. These project how global GHG emissions evolve under a successful implementation of the INDCs. These projections are based on ten global INDC analyses^{19–29} (Supplementary Table 2 provides an overview), in which calculations can be based on official estimates from countries or on documents submitted to the UNFCCC (such as national GHG inventories, national communications, biennial reports or biennial update reports). INDCs were submitted before the Paris summit; under the Paris Agreement, future mitigation contributions will be referred to as NDCs, without the ‘intended’.

2°C scenarios. These are idealized global scenarios limiting warming to well below 2°C, keeping open the option of strengthening the global temperature target to 1.5°C. These scenarios are based on a subset of scenarios from the IPCC AR5 Scenario Database (Supplementary Table 3) that meet the following criteria: they have a greater than 66% probability of keeping warming to below 2°C by 2100 (this probability does not drop below 60% at any point during the entire twenty-first century); until 2020, they assume that actions that were pledged earlier under the UNFCCC Cancun Agreement are fully implemented; and, after 2020, they distribute emission reductions across regions, gases and sectors so that the total discounted costs of the necessary global reductions are minimized. These scenarios distribute emissions reductions among regions in the most cost-optimal way, and are often referred to as least-cost or cost-optimal trajectories. However, this does not imply that the actual costs to achieve this cannot be distributed differently, for example, on the basis of other equity principles⁷⁹.

A separate set of scenarios is used to examine the post-2030 implications of current INDCs for 2°C (Supplementary Table 4).

All scenarios are expressed in terms of billions of tonnes of global annual CO₂-equivalent emissions (Gt CO₂-eq yr⁻¹). CO₂ equivalence of GHGs has been calculated by means of 100-year global warming potentials⁸⁰.

further reduce global GHG emissions in 2030 from their INDC levels towards levels that are more consistent with a long-term global pathway that limits warming to well below 2°C.

We use four scenario groups to frame the implications of the INDCs for global GHG emissions in 2030: no-policy baseline scenarios, current-policy scenarios, INDC scenarios and least-cost 2°C scenarios. Their definitions and descriptions are provided in Box 1.

Aggregate emissions impact of INDCs

A first, obvious question to ask is what the submitted INDCs deliver in terms of GHG emissions out to 2030. What sounds like simple arithmetic turns out to be a more complicated accounting exercise with an array of possible outcomes. Some countries provide a range instead of a single number of emissions reductions in their INDCs. Many INDCs lack necessary details, such as clarity on sectors and gases covered, details on the impact of listed mitigation actions, different metrics to aggregate gases, details on base year or reference values from which reductions or improvements would be measured, or accounting practices related to land use and the use of specific market mechanisms³¹. This murkiness complicates a precise estimate of their impact on emissions. Finally, some of the actions listed in INDCs are, either implicitly or explicitly, conditional on other factors, such as the availability of financial or technological support. All these factors can be interpreted differently and influence the range of possible outcomes. In our assessment, we distinguish between a conditional and an unconditional INDC scenario, with associated uncertainties. Interestingly, the Paris Agreement does not adopt such distinction, and instead defers any discussion on features of countries’ contributions to further negotiations.

Unconditionally, the INDCs are expected to result in global GHG emissions of about 55 (52–58; 10th–90th percentile range over all studies unless otherwise stated) billion metric tonnes of annual CO₂-equivalent

emissions (Gt CO₂-eq yr⁻¹; Box 1, Fig. 1, Supplementary Text 1) in 2030. This is a reduction of around 9 (7–13) Gt CO₂-eq yr⁻¹ by 2030 relative to the median no-policy baseline scenario estimate and of around 4 (2–8) Gt CO₂-eq yr⁻¹ relative to the median current-policy scenario estimate (Supplementary Table 5). Putting this into context, global GHG emissions in 2010 are estimated at about 48 Gt CO₂-eq yr⁻¹ (46–50 Gt CO₂-eq yr⁻¹; range across studies, Supplementary Table 2), and our median no-policy baseline estimate reaches about 65 Gt CO₂-eq yr⁻¹ by 2030.

A number of countries place conditions—for example, the provision of international finance—on all or part of their INDC. Some countries (such as Mexico, Indonesia and Morocco) included a range of reduction targets in their INDC and attach conditions to the implementation of the more ambitious end. Other countries indicate that their entire INDC is conditional. Of the INDCs submitted by 12 December 2015, roughly 45% came with both conditional and unconditional components; about a third was conditional only; and the remainder did not specify conditions³². When we assume in our evaluation that all conditions are met and conditional INDCs are fully implemented, estimated global GHG emissions end up about 2.4 (1.2–4.8) Gt CO₂-eq yr⁻¹ lower in 2030 compared to the unconditional INDC scenario case (full range across six available estimates, Supplementary Text 1, Supplementary Table 5).

Comparing the INDC scenario (what countries propose as their contribution to the international agreement) to the current-policy scenario (what countries implement domestically) provides lessons on the extent to which additional national policies are necessary to achieve the intended 2030 emissions reductions¹⁸. Projected emissions under current policies that match (or are lower than) those under the INDC can result either from a proactive and coordinated domestic policy response consistent with the INDC or from an INDC that is explicitly designed not to require

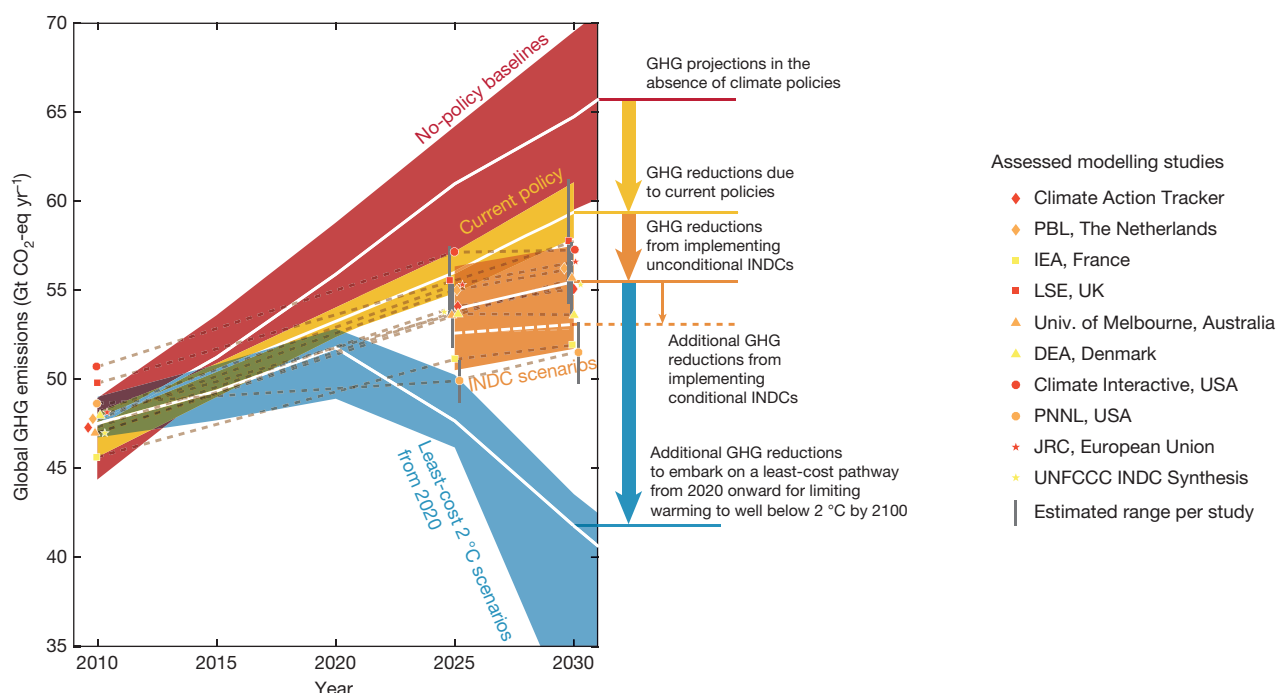


Figure 1 | Global greenhouse gas emissions as implied by INDCs compared to no-policy baseline, current-policy and 2°C scenarios. White lines show the median of each range. The white dashed line shows the median estimate of what the INDCs would deliver if all conditions are met. The 20th–80th-percentile ranges are shown for the no-policy baseline and 2°C scenarios. For current-policy and INDC scenarios, the

minimum–maximum and 10th–90th-percentile range across all assessed studies are given, respectively. Symbols represent single studies, and are offset slightly to increase readability. Dashed brown lines connect data points for each study. References to all assessed studies are provided in Box 1. Scenarios are also described in Box 1.

further policy effort. Likewise, projected emissions under current policy that exceed those under the INDC can result from a relatively ambitious INDC, from a lack of domestic climate policy, or a combination thereof. Therefore, this comparison alone cannot adequately reflect the overall level of ambition.

For a number of countries (such as Russia and Ukraine), the INDC targets suggest that emission levels above their estimated no-policy baseline or current-policy scenario will be reached. These countries are thus expected to overachieve their INDC targets by default. Under the rules of the Kyoto Protocol, over-delivery on a target would have generated surplus emission allowances by the quantity the target level is overachieved. These allowances can then be traded with other countries, who apply them to achieve their own GHG reduction target. Such a system could also be developed under the Paris Agreement, which allows for the voluntary use of “internationally transferred mitigation outcomes”. However, the extent to which such a mechanism will ultimately be developed and used remains unclear, because it will require features, information and accounting of contributions to become much more precise than they are now. Different modelling teams treat these surpluses in different ways, which adds an uncertainty of about 1 Gt CO₂-eq yr⁻¹ to the estimates presented here.

Confounding factors

The literature synthesized in this assessment reveals a wide range of estimates of future emissions under nominally similar scenarios (see small symbols in Fig. 1). These differences can stem from a number of factors, including modelling methods, input data and assumptions regarding country intent. Our review identifies four key factors that contribute to the discrepancies and differences between the various 2030 emissions estimates.

Incomplete coverage

Several global and national sectors as well as countries are not covered by INDCs. Often, emissions estimates for sectors that are not included under INDCs range widely. This is the case for, for example, global

emissions from international aviation (despite an industry pledge outside the UNFCCC³³) and maritime transport, or the national non-CO₂ GHG emissions from China. Subtracting national sectors that are not covered, INDCs cover at least 8 percentage points less of global emissions than the 96% indicated earlier (Supplementary Text 2). Under the Paris Agreement, developing countries are encouraged to move over time to economy-wide targets, so that future analyses should become more comprehensive. Countries that are not a UNFCCC Party or have not yet put forward an INDC are also studied in less depth, but represent only a diminishing amount of global emissions (about 1%–2%). Finally, studies themselves make specific choices about which INDCs to cover or focus on, which in turn influence projected emissions.

Uncertain projections

GHG emission projections of countries that have submitted INDCs are uncertain, particularly if targets are not unambiguously translatable in absolute emission reductions. Most INDCs do define straight-forward, absolute GHG emission targets (in units of CO₂-eq in a given year or period), or targets that can be relatively easily translated into absolute levels (for example, a reduction from a fixed historical base year), but this is not always the case. About 75 INDCs are defined relative to hypothetical ‘business-as-usual’ or reference scenarios in the absence of climate policy³². In some cases governments do not define their reference scenario, and in other cases official projections differ substantially from those from international and national modelling teams. Overall, these uncertainties should become smaller, because the Paris decisions request countries to ensure some methodological consistency of future submissions. Another complicating factor is that several countries put forward targets that do not directly specify emissions (such as a renewable energy target) or targets on emissions intensity (for instance, improvements of the ratio of carbon emissions, CO₂, to economic output, GDP). If the expected GDP growth rate is not provided, additional assumptions are required to quantify the implied absolute level of GHG emissions and these assumptions differ across modelling groups. For example, the estimated emissions for China for 2030 under its INDC range from 12.8 Gt CO₂-eq yr⁻¹

Table 1 | Estimates of global temperature rise for INDC and other scenarios categories

Scenario	Global-mean temperature rise by 2100 (in °C) that is not exceeded with the given probability		
	50%	66%	90%
No-policy baseline	4.1 (3.5–4.5) [3.1–4.8]	4.5 (3.9–5.1) [3.4–5.4]	5.6 (4.8–6.3) [4.2–6.8]
Current policy	3.2 (3.1–3.4) [2.7–3.8]	3.6 (3.4–3.7) [2.9–4.1]	4.4 (4.2–4.6) [3.6–5.2]
INDC (unconditional)	2.9 (2.6–3.1) [2.2–3.5]	3.2 (2.9–3.4) [2.4–3.8]	3.9 (3.5–4.2) [2.8–4.7]
INDC (conditional)	2.7 (2.5–2.9) [2.1–3.2]	3.0 (2.7–3.1) [2.2–3.6]	3.7 (3.3–3.9) [2.6–4.4]

For each scenario, temperature values at the 50%, 66% and 90% probability levels are provided for the median emission estimates, as well as the 10th–90th-percentile range of emissions estimates (in parentheses) and the same estimates when also including scenario projection uncertainty (in brackets). Temperature increases are relative to pre-industrial levels (1850–1900), and are derived from simulations with a probabilistic set-up with the simple model MAGICC (refs 10, 68–70, Supplementary Text 3).

to 15.0 Gt CO₂-eq yr⁻¹ in different studies^{22,34,35}. At least seven other INDCs, including India's, are subject to the same kind of uncertainties (Supplementary Table 1). Finally, many countries (about 30, amounting to approximately 6% of global emissions) include mere qualitative descriptions of mitigation actions in their INDCs, which complicate a precise quantification.

Land-use-related emissions

Various approaches exist to account for emissions from land use, land-use change and forestry, and countries can use an accounting approach of their choice in their INDCs. Examples of possible approaches are to include land-use-related CO₂ emissions and removals as part of the national total, much like any other sector (an approach favoured by, for example, Brazil and USA), or to apply accounting rules similar to the ones under the Kyoto Protocol (which are favoured by, for example, the European Union and New Zealand, and possibly by Russia). These accounting rules can have a substantial effect on the emissions of individual countries in 2025 and 2030³⁶ and are associated with substantial uncertainties. Although some INDCs explicitly exclude land-use-related emissions from their targets, many INDCs that include land use in their targets do not specify an accounting approach.

Historical emissions and metrics

Historical emission estimates come with their associated uncertainties. For example, recently, global 2010 GHG emissions have been estimated at 49 Gt CO₂-eq yr⁻¹ (±4.5 Gt CO₂-eq yr⁻¹, 90% confidence interval)³⁷. Model teams apply their own estimate of historical emissions in their INDC analyses (Fig. 1), and both INDCs and analysts use varying metrics to translate GHG emissions into units of CO₂-equivalence. Even if these discrepancies can be harmonized³⁸ or corrected for, their variation increases the uncertainty surrounding INDC estimates.

Optimal 2°C pathways

Having quantified the GHG implications of the INDCs by 2030, the question remains whether these levels are consistent with the Paris Agreement's aim of holding warming to well below 2°C. As indicated earlier, limiting warming to any level requires net CO₂ emissions to become zero at some point in time and, given the small remaining carbon budget, this moment is estimated to be before the end of this century for a 2°C limit^{11,17}. The Paris Agreement's aim of reaching net-zero GHG emissions in the second half of the century goes even further. For some non-CO₂ emissions, in particular those related to agriculture, only limited mitigation options have been identified³⁹. Therefore, net-zero CO₂ emissions are always achieved before achieving net-zero GHG emissions. Integrated energy–economy models are used extensively to model pathways that can achieve this feat at global least cost¹⁶. Here, we use the Scenario Database that accompanied the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) to explore such cost-optimal 2°C pathways from 2020 onward (Box 1).

Comparing these cost-optimal 2°C scenarios to the INDC projections shows a large discrepancy (Fig. 1). The median cost-optimal path towards keeping warming to below 2°C (starting reductions in 2020) and the emissions currently implied by the unconditional INDCs differ by about 14 (10–16) Gt CO₂-eq yr⁻¹ in 2030. Even if the conditions that are linked

to some INDCs are met (see earlier), this difference remains of the order of 11 Gt CO₂-eq yr⁻¹. The high end of this range (16 Gt CO₂-eq yr⁻¹) corresponds roughly to the 2010 emissions of China and USA combined; the lower end (about 10.5 Gt CO₂-eq yr⁻¹) to the sum of the emissions of Brazil, the European Union, India and Russia. Thus, the INDCs clearly do not put the world on a least-cost path towards limiting warming to well below 2°C.

Any global emission scenario reflects an idealized representation of the world. This is not different for the cost-optimal 2°C scenarios that were used above as a reference. The strength of such cost-optimal scenarios lies in the fact that they provide an assessment of the potential for emission reductions in a world that collaborates globally towards limiting climate change and attempts to do this at lowest overall cost. Other scenarios in the literature model other, more imperfect futures, for example, those in which climate action is delayed by a few decades^{13,14,40,41}, in which countries and regions are not collaborating from the beginning⁴², or in which the strength of local institutions affects the willingness to invest⁴³. Such scenarios help us to explore the post-2030 implications of the current INDCs.

Post-2030 implications of INDCs

A large share of the potential warming until 2100 is determined not just by the INDCs until 2025 or 2030, but also by what happens afterwards. Several conceptual approaches can be followed to extend INDCs into the future, which basically assume that climate action stalls, continues or accelerates. Stalling action is often modelled by assuming that emissions return to a no-climate-policy trajectory after 2030; continuing action by assuming that the level of post-2030 action is similar to pre-2030 action on the basis of a metric of choice (for example, extrapolating INDC trends in terms of carbon-price development or emissions intensity of the economy); and accelerating action by post-2030 action that goes beyond such a level. Because of the path-dependence and inertia of the global energy system^{44,45}, the INDCs have a critical role in preparing what can come after 2030.

Each of the above-mentioned approaches leads to different global temperature outcomes, even when starting from the same INDC assessment for 2030. It is therefore essential to spell out post-2030 assumptions to understand global temperature projections for the twenty-first century based on the INDCs. As a conservative interpretation of the Paris Agreement, we here assume that climate action continues after 2030 at a level of ambition that is similar to that of the INDCs (Supplementary Text 3). The assumption that climate action will continue or accelerate over time is supported by the Agreement's requirement that the successive nationally determined contribution of each country must represent a progression beyond the earlier contributions, and reflect the highest possible ambition, of that country. Stalling climate action after 2030 would be in contradiction with the provisions of the Paris Agreement.

Under these assumptions of continued climate action, the 2030 unconditional-INDC emission range is roughly consistent with a median warming relative to pre-industrial levels of 2.6–3.1°C (median, 2.9°C; full scenario projection uncertainty, 2.2–3.5°C; Table 1, Fig. 2b, Supplementary Fig. 1), with warming continuing its increase afterwards. This is an improvement on the current-policy and no-policy baseline scenarios, whose median projections suggest about 3.2°C and more than 4°C of temperature rise by 2100, respectively. The successful implementation

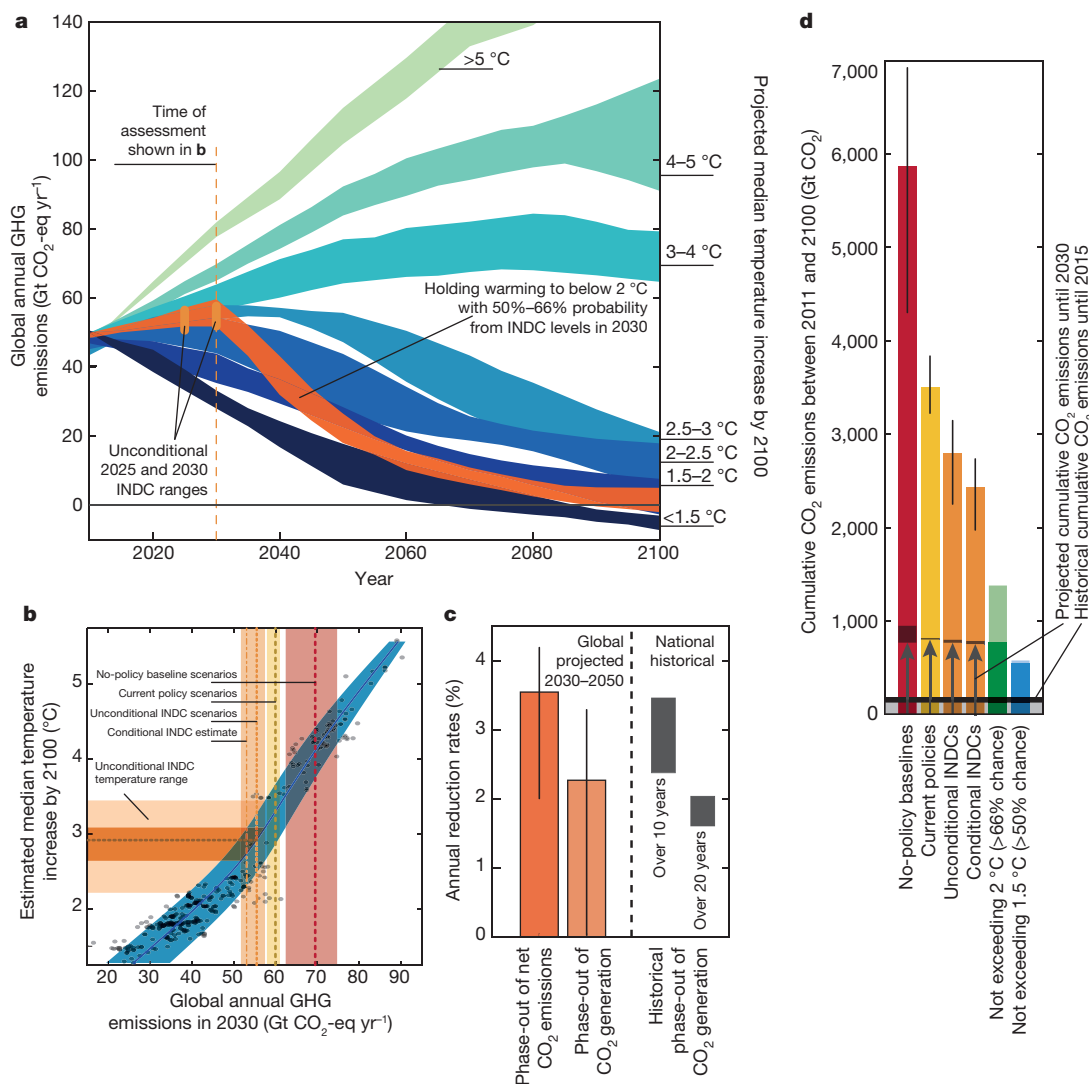


Figure 2 | Temperature implications of current INDCs. **a**, GHG emission ranges (20th–80th percentile) of scenarios from the IPCC AR5 Scenario Database with constant policy assumptions from 2010 onwards (blue-to-green shaded ranges), grouped per estimated median global-mean temperature increase in 2100 relative to pre-industrial levels (1850–1900), and range of the scenario subset limiting warming to below 2 °C by 2100 with 50%–66% likelihood (dark orange) from year-2030 INDC levels. The vertical orange lines show the unconditional INDC range in 2025 and 2030, as shown in Fig. 1. The 2 °C range shown in Fig. 1 starts global least-cost mitigation action in 2020 instead of 2010 and is not included here. **b**, Relationship between global GHG emission levels in 2030 and median global-mean temperature increase by 2100 based on scenarios shown in **a**. Each dot represents a single scenario. The blue line shows a smoothing spline fit ($R^2 \approx 0.93$) and the blue-shaded area shows fits to the 5th and 95th percentile over all points. Comparing the central fit with the range of year-2030 GHG emissions implied by the unconditional INDCs shows that INDCs are roughly consistent with a median warming of

2.6–3.1 °C by 2100 (horizontal dark-orange range), and a 2.2–3.5 °C range including scenario projection uncertainty (horizontal light-orange range). Vertical dashed lines and shaded regions show year-2030 GHG estimates for the various scenario sets. **c**, Annual CO₂ reduction rates modelled in scenarios limiting warming to below 2 °C from year-2030 INDC levels (dark-orange range in **a**; bars, median; vertical lines, spread across all available scenarios) and historical examples (range for France, Sweden and Denmark is based on ref. 74; see Supplementary Text 4). **d**, Implied cumulative carbon emissions including uncertainties, and comparison to budget ranges for not exceeding 1.5 °C (with 50% probability) and 2 °C (with 66% probability) from refs 9 and 11 (dark bar, lower estimate; light bar, high-range estimate). Historical estimates are from ref. 75. Vertical lines show the range due to scenario spread (Supplementary Text 3 and Supplementary Table 6). Arrows and bars in the first four columns show the projected cumulative CO₂ emissions until 2030 for each respective scenario.

of all conditional INDCs would decrease our median estimate by an additional 0.2 °C, but keeps the outcome far from the world the Paris Agreement is aiming for, with well-below 2 °C and 1.5 °C of warming. Moreover, all above-mentioned values represent median projections. Because the climate response to GHG emissions remains uncertain⁴⁶, it is also possible that substantially higher temperatures will materialize with compelling likelihoods (Table 1). For example, at the 66th percentile level, warming under the unconditional INDCs is projected to be about 0.3 °C higher (3.2 °C, with a range of 2.9–3.4 °C). Finally, the INDC cases that we assess here will exceed the available carbon budget for keeping warming to below 2 °C by 2030 with 66% probability (that is, roughly

750–800 Gt CO₂ implied emissions under the INDCs during the 2011–2030 period compared to the 750–1,400 Gt CO₂ available; Supplementary Text 3, Supplementary Table 6, Fig. 2d). The budget for never exceeding 1.5 °C with a 50% probability (550–600 Gt CO₂) will be entirely gone, indicating that active removal of CO₂ at a later point in time will be required to return to within this budget (Supplementary Table 6). Median warming under the INDCs is projected to cross the 1.5 °C and 2 °C limits by 2030–2045 and 2045–2075, respectively (Supplementary Fig. 4).

The question thus arises whether global temperature rise can be kept to well below 2 °C with accelerated action after 2030. Global scenarios that aim to keep warming to below 2 °C and that achieve this objective

from 2030 GHG emissions similar to those from the INDC range have been assessed in detail by recent large-scale model-comparison projects^{16,40}. Our re-analysis of these scenarios shows that even with accelerated action after 2030, options to keep warming to well below 2 °C from current INDCs are severely limited, particularly if some key mitigation technologies do not scale up as anticipated. This is easy to understand if one appreciates that even if all INDCs are successfully implemented by 2030, the 2 °C carbon budget might already be virtually exhausted by that time (see earlier and Fig. 2d). The Paris Agreement does not define precisely what its “well-below 2 °C” aim means. Typically, policymakers in the UNFCCC have been concerned about limiting warming to below 2 °C with >66% probability⁴⁷. However, from current INDC levels, all available internally consistent scenarios manage to limit warming to below 2 °C with only a lower, 50%–66% probability, increasing the risks of climate change impacts. No scenarios are available that are consistent with both the current INDCs and a 1.5 °C warming limit with 50% probability.

The available scenarios show rapidly declining emissions after 2030, with global CO₂ emissions from energy- and industry-related sources reaching net-zero levels between 2060 and 2080. The global economy is thus assumed to fully decarbonize¹⁷ in the time span of three to five decades and from 2030 levels that are higher than today's. Furthermore, about two thirds of these scenarios achieve a balance of global GHG emissions (as mentioned in the Paris Agreement) between 2080 and 2100. Because some non-CO₂ emissions are virtually impossible to eliminate entirely (for example, those from specific agricultural sources³⁹), reaching such a balance will involve net-negative CO₂ emissions⁴⁸ at a global scale to compensate for any residual non-CO₂ emissions¹⁷, resulting in a gradual decline in global-average temperatures over time. Technologies that might be able to achieve this feat are still surrounded by important uncertainties (see below). In general, lower near-term emissions allow for a later timing of reaching global net-zero CO₂ emissions¹⁷ (see 1.5–2 °C versus dark-orange range in Fig. 2a) and, moreover, reduce the overall future reliance on negative emissions technologies^{16,40,41}.

To illustrate the challenges involved, we take a critical look at some characteristics of the scenarios. Scenarios that broadly follow the INDCs until 2030 and still manage to keep warming to below 2 °C (with 50%–66% probability only) are associated with a very rapid decline in CO₂ emissions from energy- and industry-related sources after 2030. The decarbonization between 2030 and 2050 is particularly decisive in these scenarios^{40,45}. For this period, the scenarios show average rates of decline in annual emissions of about 3.5% (2.0%–4.2%, full range across scenarios; Supplementary Text 4). To understand what this means in a historical context, it makes sense to distinguish between (1) the phase-out of CO₂ generation over time (a proxy for the reduction in fossil-fuel use and upscaling of low-carbon energy sources) and (2) the required upscaling of industrial-scale CO₂ sequestration with carbon capture and geological storage (CCS) technologies^{49,50}. The latter mitigation option has not been applied in the past. It can thus be seen as an additional technological option that is included in scenarios, but that did not contribute to the past experiences.

In the 2 °C scenarios that start from INDC levels in 2030 (dark-orange range in Fig. 2a), CO₂ generation is reduced at a median annual rate of about 2.3% (0.0%–3.3%, full range) between 2030 and 2050. Historically, countries have been able to achieve reductions in CO₂ generation at rates of about 2%–3% per year as a result of dedicated (energy-security) policies⁴⁰ (Fig. 2c; Supplementary Text 4). Limiting warming to below 2 °C from year-2030 INDC levels thus implies that the pace of such a precipitated phase-out of fossil-fuel use needs to be replicated globally. These historical reductions were all achieved for non-climate reasons, with a focus on energy security and not on emissions reductions. There is thus no clear historical analogue for reductions under a dedicated and stringent climate policy. The challenge remains nevertheless important. This becomes even clearer when appreciating that all historical analogues for reductions were achieved in highly developed countries, such as France, Sweden and Denmark. Achieving similar results in developing countries, with energy-intensive sectors that are still growing and with weaker

institutional frameworks, higher investment risks and less capacity, will be more difficult⁴³, but, at the same time, readily available low-cost zero-carbon alternatives could also allow those economies to leap-frog carbon-intensive development in some sectors.

Scenarios complement the global phase-out of CO₂ generation with a scale-up of CCS infrastructure to capture and geologically store part of the CO₂ that continues to be generated. This scale-up is massive in scenarios that limit warming to below 2 °C from INDC levels. Because such scenarios have limited CCS deployment until 2030, the annual rate of CO₂ sequestration is assumed to increase 10- to >100-fold in the 2030–2050 period, reaching about 10 Gt CO₂ yr^{−1} in 2050 (8–14 Gt CO₂ yr^{−1} range). To put this challenge into perspective, about 85 GW (measured in coal-equivalent power generation; Supplementary Text 4) of new CCS capacity would need to be installed each year to capture this amount by 2050, which corresponds roughly to the combined capacity of solar and wind power generation that is annually globally installed today^{51,52} (Supplementary Fig. 2). Altogether, the global energy-system transition that is required to limit warming to well below 2 °C and further to 1.5 °C is unprecedented.

Finally, scenarios often combine CCS with biomass energy (abbreviated as BECCS) as a way of actively capturing and removing CO₂ from the atmosphere. Although in principle this is technically possible, deployment of such technologies at scale is untested, and could be controversial because of public acceptance⁵³ or because of their competition with food production over land and water^{53,54}. A recent review⁵⁴ showed that, assuming agricultural practices and yields do not change over the twenty-first century, removing CO₂ could require large amounts of land. At the same time, other assessments^{39,55–57} concluded that it might be possible to produce the required amount of bio-energy in a sustainable way (up to 300 EJ yr^{−1}, see box 11.5 in ref. 39). The importance of the land-use question for policy is highlighted by the decision of the IPCC to dedicate one of its three upcoming Special Reports to questions of sustainable land management and food security. Exploring futures in which a global balance of GHG emissions can be achieved in the second half of this century with technically feasible and societally acceptable technologies represents a major research challenge emerging from the Paris Agreement⁵⁸. This challenge is particularly relevant to policy, because limiting emissions in 2030 does not only increase the chances of attaining the 2 °C target, but also reduces the need to rely on unproven, potentially risky or controversial technologies in the future^{16,40,59}.

Decreasing the post-2030 challenge

The post-2030 challenge to limit warming to below 2 °C from current INDC levels is daunting, and pursuing efforts for 1.5 °C even more so. However, the overall challenge can be minimized by additional GHG reductions in the near-term^{16,40,41}. In this context, near-term means before and by 2030. Besides (i) the option of countries increasing the overall ambition of their INDCs, we identify several other options that can contribute to this (see Table 2 for an overview). The options include: (ii) increasing the coverage of INDCs to more sectors and gases; (iii) including international sectors such as aviation and international maritime transport; (iv) implementing measures that enable over-delivery on the INDCs; (v) increasing contributions to international climate finance and international cooperation on technology development, transfer and diffusion; and (vi) promoting and implementing additional national, sub-national and non-state initiatives. These options are not fully additional; some of them overlap (strongly) with the INDCs, and their precise contributions thus remain speculative (see Table 2). However, several indications suggest that such an increase in ambition is possible.

First, increasing ambition over time is a key component of the Paris Agreement framework. For example, countries are requested to submit new—or update existing—contributions that should represent a progression beyond their earlier commitments. The certainty of the new global climate agreement, together with the improving cost and availability of low-carbon technologies⁶⁰, might help countries to consider

Table 2 | Overview of options to reduce post-2030 challenge

Option	Description	Possible impact on global emissions in 2030 (order of magnitude)
(i) Increasing ambition of existing 2025 and 2030 contributions	The outcome of the Paris climate summit provides several opportunities to increase ambition of national contributions by 2030, for example, through consecutive five-year cycles during which national contributions increase in ambition.	10 Gt CO ₂ -eq yr ⁻¹ Theoretical potential to embark on a least-cost 2 °C pathway after 2020 ¹⁶ .
(ii) Increasing coverage of sectors and gases	Some countries cover only part of their total GHG emissions and some sectors in their contributions; for example, some contributions apply only to CO ₂ and not to other GHGs. Extending INDCs to all sectors and gases would increase the global coverage of INDCs.	0.1–1 Gt CO ₂ -eq yr ^{-1†}
(iii) Including international sectors	At present, the contributions cover only countries. International sectors, such as international aviation and maritime transport can also be included. These sectors covered around 2% of global emissions in 2010 [‡] , with an increasing trend.	0.1–1 Gt CO ₂ -eq yr ⁻¹ (ref. 71)
(iv) Implementing domestic measures that enable over-delivery on the INDCs*	Countries can implement domestic measures that go beyond the actions described in the current INDCs.	10 Gt CO ₂ -eq yr ⁻¹ Theoretical potential to embark on a least-cost 2 °C pathway ^{16,72} .
(v) Increasing climate finance and international cooperation*	Additional international climate finance and cooperation on technology development, transfer and diffusion could help to (over-)achieve the conditional end of the national contributions.	1 Gt CO ₂ -eq yr ⁻¹ Estimate for moving from unconditional to conditional INDCs [§] . No estimate available for additional reductions.
(vi) Implementing international cooperative initiatives*	Action could be implemented by ambitious sub-national or regional governments, companies, organizations, non-governmental organisations and citizens to further reduce emissions. The amount of overlap of these initiatives with national contributions remains unclear.	1 Gt CO ₂ -eq yr ⁻¹ in 2020 No comprehensive estimates available for 2030.

Further reducing global GHG emissions by 2030 reduces the post-2030 challenge for limiting global-mean temperature increase to below 2 °C relative to pre-industrial levels. This table provides an overview of options to improve pre-2030 emissions reductions. These options can overlap to smaller or larger degrees with the nationally determined contributions indicated in option (i); consequently, they are not fully additive. Estimates of the potential impact on global emissions are highly uncertain and therefore only the expected order of magnitude (nearest power of 10) is provided for each option.

*Options that overlap to smaller or larger degrees with option (i).

†See Supplementary Text 2.

‡Source: Emission Database for Global Atmospheric Research (EDGAR), <http://edgar.jrc.ec.europa.eu/>.

§See Supplementary Table 5.

||Estimate based on refs 64, 65 and 73.

strengthening their post-2020 contributions. Second, countries can undertake further domestic measures. Because many countries have undergone national stakeholder processes in preparation of their INDCs, they could now be in a better position to consider additional policies⁶¹. Sub-national actors such as cities and regional governments may take further action, and non-state actors can also help to overachieve INDCs. The Paris conference saw unprecedented willingness to act by these stakeholders, with more than 1,000 non-state actors signing the Paris Pledge for Action (<http://www.parispledgeforaction.org/>), signalling that they are willing to support efforts to meet and exceed the ambition of governments for keeping the world on a 2 °C trajectory. This role of non-Party stakeholders is acknowledged more clearly than ever before in the official Paris decisions. However, although the theoretical potential of these activities is huge, their additional impact is very hard to quantify, and it remains unclear whether these initiatives are additional to the already pledged national contributions^{62–65}.

Outlook

Covering most of the world's GHG emissions with climate plans in the form of voluntarily submitted INDCs is a historic achievement. The Paris Agreement requires the submission of successive, increasingly ambitious, nationally determined contributions that are subject to strong transparency guidelines, as well as a global stock-take, in the light of equity and science, every five years. The optimism accompanying this process has to be carefully balanced against the important challenges that current INDCs imply for post-2030 emissions reductions. Even starting today, limiting warming to no more than 2 °C relative to pre-industrial levels constitutes a societal challenge; at the same time, the warming projected from current INDCs constitutes an important challenge on its own in terms of coping with climate impacts. The nationally determined contributions constitute a new era for climate policy under the Paris Agreement, and represent both an invitation and a call for further action. Furthering greater reductions in the coming decade and preparing for a global transformation of development pathways is critical.

Two developments look particularly promising to us. First, it becomes increasingly clear to decision-makers that measures to reduce GHG emissions have multiple socio-economic benefits⁶⁶. The action by virtually all countries improves prospects for further collective action, which must be the fundamental basis of any adequate response to climate change. Therefore, it becomes easier to conceive additional measures or strengthen existing ones. Second, the recent unprecedented engagement of non-state actors such as businesses, citizens and religious organizations illustrates a more profuse awareness and an increased momentum for climate action. Given the large potential for emissions reductions as a result of both of these options, supporting and enabling national and non-state action will be critical. This insight also opens important avenues for future research and assessment. The research community will have to break from a one-sided climate-policy-centred approach and develop new concepts and frameworks that further the achievement of a portfolio of societal objectives, including climate, food and energy security, public health, and other goals of the sustainable development agenda⁶⁷. Charting development pathways that can hold warming well below 2 °C will thus require a renewed effort of the social and physical science communities alike.

Received 1 September 2015; accepted 9 May 2016.

1. UNFCCC. *Adoption of the Paris Agreement*. Report No. FCCC/CP/2015/L.9/Rev.1, <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf> (UNFCCC, 2015).
2. UNFCCC. *INDCs as communicated by Parties*. <http://www4.unfccc.int/submissions/indc/Submission%20Pages/submissions.aspx> (2015).
3. UNFCCC. *United Nations Framework Convention on Climate Change*. Report No. FCCC/INFORMAL/84, <https://unfccc.int/resource/docs/convkp/conveng.pdf> (UNFCCC, 1992).
4. Knutti, R., Rogelj, J., Sedláček, J. & Fischer, E. M. A scientific critique of the two-degree climate change target. *Nat. Geosci.* **9**, 13–18 (2016).
5. Randalls, S. History of the 2 °C climate target. *Wiley Interdiscip. Rev. Clim. Change* **1**, 598–605 (2010).
6. Collins, M. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the*

- Intergovernmental Panel on Climate Change* (eds Stocker T. F. et al.) 1029–1136 (Cambridge Univ. Press, 2013).
7. Matthews, H. D. & Caldeira, K. Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, L04705 (2008).
 8. Knutti, R. & Rogelj, J. The legacy of our CO₂ emissions: a clash of scientific facts, politics and ethics. *Clim. Change* **133**, 361–373 (2015).
 9. IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2014).
- The most recent intergovernmental report synthesizing our current understanding of the physical science and impacts of, and possible solutions to avoid, climate change.**
10. Meinshausen, M. et al. Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
 11. Rogelj, J. et al. Differences between carbon budget estimates unravelled. *Nat. Clim. Change* **6**, 245–252 (2016).
 12. Friedlingstein, P. et al. Persistent growth of CO₂ emissions and implications for reaching climate targets. *Nat. Geosci.* **7**, 709–715 (2014).
 13. Rogelj, J., McCollum, D. L., Reisinger, A., Meinshausen, M. & Riahi, K. Probabilistic cost estimates for climate change mitigation. *Nature* **493**, 79–83 (2013).
 14. Luderer, G. et al. Economic mitigation challenges: how further delay closes the door for achieving climate targets. *Environ. Res. Lett.* **8**, 034033 (2013).
 15. Rogelj, J. et al. Energy system transformations for limiting end-of-century warming to below 1.5 °C. *Nat. Clim. Change* **5**, 519–527 (2015).
 16. Clarke, L. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer O. et al.) 413–510 (Cambridge Univ. Press, 2014).
 17. Rogelj, J. et al. Zero emission targets as long-term global goals for climate protection. *Environ. Res. Lett.* **10**, 105007 (2015).
- Study providing insights about the consistency between Article 2 and Article 4 of the UNFCCC Paris Agreement.**
18. UNEP. *The Emissions Gap Report 2015* http://uneplive.unep.org/media/docs/theme/13/EGR_2015_301115_lores.pdf (UNEP, 2015).
- An annual assessment tracking climate policy action over the past six years, which provided the basis for the analysis presented in this Perspective.**
19. Climate Analytics, Ecofys, NewClimate Institute & PIK. Climate pledges will bring 2.7°C of warming, potential for more action. *Climate Action Tracker* <http://climateactiontracker.org/news/253/Climate-pledges-will-bring-2.7C-of-warming-potential-for-more-action.html> (2015).
 20. Admiraal, A. et al. *Assessing Intended Nationally Determined Contributions to the Paris Climate Agreement – What are the Projected Global and National Emission Levels for 2025–2030?* Report No. PBL 1879, <http://www.pbl.nl/en/publications/assessing-intended-nationally-determined-contributions-to-the-paris-climate-agreement> (PBL Netherlands Environmental Assessment Agency, 2015).
 21. IEA. *World Energy Outlook 2015. International Energy Agency* <http://www.worldenergyoutlook.org/resources/energyandclimatechange/> (2015).
 22. Boyd, R., Cranston Turner, J. & Ward, B. *Intended Nationally Determined Contributions: What are the Implications for Greenhouse Gas Emissions in 2030?* http://www.lse.ac.uk/GranthamInstitute/wp-content/uploads/2015/10/Boyd_Turner_and_Ward_policy_paper_October_2015.pdf (Centre for Climate Change Economics and Policy, and Grantham Research Institute on Climate Change and the Environment, 2015).
 23. Meinshausen, M. INDC Factsheets. *Australian-German Climate and Energy College* <http://climate-energy-college.net/indc-factsheets> (2015).
 24. DEA. *Analyzing the 2030 emissions gap. Danish Energy Agency* <http://www.ens.dk/gap> (2015).
 25. Climate Interactive. *Climate Scoreboard. Climate Interactive* <https://www.climateinteractive.org/programs/scoreboard/> (2015).
 26. Fawcett, A. A. et al. Can Paris pledges avert severe climate change? *Science* **350**, 1168–1169 (2015).
 27. UNFCCC. *Synthesis Report on the Aggregate Effect of the Intended Nationally Determined Contributions*. Report No. FCCC/CP/2015/7, <http://unfccc.int/resource/docs/2015/cop21/eng/07.pdf> (UNFCCC, 2015).
 28. Kitous A. & Keramidas K. *Analysis of Scenarios Integrating the INDCs* <https://ec.europa.eu/jrc/sites/default/files/JRC97845.pdf> (Joint Research Centre, 2015).
 29. den Elzen, M. et al. *Enhancing Mitigation Ambitions in the Major Emitting Countries: Analysis of Current and Potential Climate Policies*. Report No. PBL 1631, <http://www.pbl.nl/en/publications/enhancing-mitigation-ambitions-in-the-major-emitting-countries-analysis-of-current-and-potential-climate-policies> (PBL Netherlands Environmental Assessment Agency, 2015).
 30. JRC/PBL. *EDGAR version 4.2FT2012* <http://edgar.jrc.ec.europa.eu/overview.php?v=GHGs1990-2012> (Joint Research Centre of the European Commission, PBL Netherlands Environmental Assessment Agency, 2014).
 31. Damassa, T. et al. *Interpreting INDCs: Assessing Transparency of Post-2020 Greenhouse Gas Emissions Targets for 8 Top-Emitting Economies (Working Paper)* http://www.wri.org/sites/default/files/WRI_WP_InterpretingINDCs.pdf (World Resources Institute, 2015).
 32. World Resources Institute. *Paris Contributions Map. CAIT Climate Data Explorer* <http://cait.wri.org/indc/> (2015).
 33. ICAO. *Assembly Resolutions in Force (as of 4 October 2013)*. Report No. 10022, http://www.icao.int/publications/documents/10022_en.pdf (ICAO, 2014).
 34. den Elzen, M. G. J. et al. Greenhouse gas emissions from current and enhanced policies of China until 2030: can emissions peak before 2030? *Energy Policy* **89**, 224–236 (2016).
 35. Sha, F., Ji, Z. & Linwei, L. *An Analysis of China's INDC* <http://www.chinacarbon.info/wp-content/uploads/2015/07/Comments-on-Chinas-INDC.pdf> (China National Center for Climate Change Strategy and International Cooperation (NCSC), 2015).
 36. Grassi G. & Dentener F. *Quantifying the Contribution of the Land Use Sector to the Paris Climate Agreement*. Report No. EUR 27561, <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC98451/jrc%20lulucf-indc%20report.pdf> (European Union, JRC Science Hub, 2015).
 37. Blanco, G. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer O. et al.) 351–411 (Cambridge Univ. Press, 2014).
 38. Rogelj, J., Hare, W., Chen, C. & Meinshausen, M. Discrepancies in historical emissions point to a wider 2020 gap between 2 °C benchmarks and aggregated national mitigation pledges. *Environ. Res. Lett.* **6**, 024002 (2011).
 39. Smith, P. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer O. et al.) 811–922 (Cambridge Univ. Press, 2014).
 40. Riahi, K. et al. Locked into Copenhagen pledges — implications of short-term emission targets for the cost and feasibility of long-term climate goals. *Technol. Forecast. Soc. Change* **90**, 8–23 (2015).
 41. Rogelj, J., McCollum, D. L., O'Neill, B. C. & Riahi, K. 2020 emissions levels required to limit warming to below 2 °C. *Nat. Clim. Change* **3**, 405–412 (2013).
 42. Kriegler, E. et al. Making or breaking climate targets: The AMPERE study on staged accession scenarios for climate policy. *Technol. Forecast. Soc. Change* **90**, 24–44 (2015).
 43. Iyer, G. C. et al. Improved representation of investment decisions in assessments of CO₂ mitigation. *Nat. Clim. Change* **5**, 436–440 (2015).
 44. Roehrl, R. A. & Riahi, K. Technology dynamics and greenhouse gas emissions mitigation: a cost assessment. *Technol. Forecast. Soc. Change* **63**, 231–261 (2000).
 45. Eom, J. et al. The impact of near-term climate policy choices on technology and emission transition pathways. *Technol. Forecast. Soc. Change* **90**, 73–88 (2015).
 46. Stocker, T. F. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker T. F. et al.) 33–115 (Cambridge Univ. Press, 2013).
 47. UNFCCC. *Establishment of an Ad Hoc Working Group on the Durban Platform for Enhanced Action*. Draft decision -/CP.17, https://unfccc.int/files/meetings/durban_nov_2011/decisions/application/pdf/cop17_durbanplatform.pdf (UNFCCC, 2011).
 48. Obersteiner, M. et al. Managing climate risk. *Science* **294**, 786–787 (2001).
 49. Scott, V., Haszeldine, R. S., Tett, S. F. B. & Oschlies, A. Fossil fuels in a trillion tonne world. *Nat. Clim. Change* **5**, 419–423 (2015).
 50. IPCC. *IPCC Special Report on Carbon Dioxide Capture and Storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change* (eds Metz, B. et al.) (Cambridge Univ. Press, 2005).
 51. Global Wind Energy Council. *Global Statistics* <http://www.gwec.net/global-figures/graphs/> [accessed 15 October 2015] (2015).
 52. SolarPower Europe. *Global Market Outlook For Solar Power / 2015 - 2019* http://helapco.gr/pdf/Global_Market_Outlook_2015_-2019_lr_v23.pdf (SolarPower Europe, 2015).
 53. Upham, P. & Roberts, T. Public perceptions of CCS: emergent themes in pan-European focus groups and implications for communications. *Int. J. Greenh. Gas Control* **5**, 1359–1367 (2011).
 54. Smith, P. et al. Biophysical and economic limits to negative CO₂ emissions. *Nat. Clim. Change* **6**, 42–50 (2016).
 55. Creutzig, F. et al. Bioenergy and climate change mitigation: an assessment. *GCB Bioenergy* **7**, 916–944 (2015).
 56. Coelho, S. T. et al. in *Global Energy Assessment: Toward a Sustainable Future* (eds Johannsson, T. B. et al.) Ch. 20, 1459–1525 (Cambridge Univ. Press, International Institute for Applied Systems Analysis, 2012).
 57. IPCC. *Special Report on Renewable Energy Sources and Climate Change Mitigation* (eds Edenhofer, O. et al.) <http://www.ipcc.ch/report/srren/> (Cambridge Univ. Press, 2011).
 58. Rogelj, J. & Knutti, R. Geosciences after Paris. *Nat. Geosci.* **9**, 187–189 (2016).
 59. Iyer, G. C. et al. The contribution of Paris to limit global warming to 2 °C. *Environ. Res. Lett.* **10**, 125002 (2015).
 60. Cronin C. et al. *Faster and Cleaner: Decarbonization in the Power and Transport Sectors is Surpassing Predictions and Offering Hope for Limiting Warming to 2 °C* <http://www.climateworks.org/wp-content/uploads/2015/11/Faster-Cleaner-Decarbonization-in-the-Power-Transport-Sectors.pdf> (ClimateWorks, NewClimate Institute, Ecofys, Climate Analytics, 2015).
 61. Day, T., Röser, F., Tewari, R., Kurdziel, M. & Höhne, N. *Preparation of Intended Nationally Determined Contributions (INDCs) as a Catalyst for National Climate Action* http://mitigationpartnership.net/sites/default/files/indc_as_catalyst.pdf (NewClimate Institute, 2015).
 62. Weischer, L., Morgan, J. & Patel, M. Climate clubs: can small groups of countries make a big difference in addressing climate change? *Rev. Eur. Community Int. Environ. Law* **21**, 177–192 (2012).
 63. Blok, K., Höhne, N., van der Leun, K. & Harrison, N. Bridging the greenhouse-gas emissions gap. *Nat. Clim. Change* **2**, 471–474 (2012).

64. Hsu, A., Moffat, A. S., Weinfurter, A. J. & Schwartz, J. D. Towards a new climate diplomacy. *Nat. Clim. Change* **5**, 501–503 (2015).
 65. UNEP. *Climate Commitments of Subnational Actors and Business: A Quantitative Assessment of their Emission Reduction Impact*. Report No. DEW/1917/NA, http://apps.unep.org/redirect.php?file=/publications/pmtdocuments/-Climate_Commitments_of_Subnational_Actors_and_Business-2015CCSA_2015.pdf.pdf (United Nations Environment Programme, 2015).
 66. IPCC in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 1–33 (Cambridge Univ. Press, 2014).
 67. UN General Assembly. *Transforming our World: the 2030 Agenda for Sustainable Development*. Report No. A/RES/70/1, http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (United Nations, 2015).
 68. Rogelj, J., Meinshausen, M., Sedláček, J. & Knutti, R. Implications of potentially lower climate sensitivity on climate projections and policy. *Environ. Res. Lett.* **9**, 031003 (2014).
 69. Rogelj, J., Meinshausen, M. & Knutti, R. Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nat. Clim. Change* **2**, 248–253 (2012).
 70. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – part 1: model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
 71. Cames, M., Graichen, J., Siemons, A. & Cook, V. *Emission Reduction Targets for International Aviation and Shipping*. Report No. IP/A/ENVI/2015-11, [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/569964/IPOL_STU\(2015\)569964_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/569964/IPOL_STU(2015)569964_EN.pdf) (European Parliament's Committee on Environment, Public Health and Food Safety, 2015).
 72. Fekete, H. et al. *The Impact of Good Practice Policies on Regional and Global Greenhouse Gas Emissions* https://newclimateinstitute.files.wordpress.com/2015/07/task2c_goodpracticeanalysis_july_2015.pdf (NewClimate Institute, PBL Netherlands Environmental Assessment Agency, International Institute for Applied Systems Analysis, 2015).
 73. Roelfsema, M., Harmsen, M., Olivier, J. & Hof, A. F. *Climate Action Outside the UNFCCC*. Report No. PBL 1188, http://www.pbl.nl/sites/default/files/cms/pbl-2015-climate-action-outside-the-unfccc_01188.pdf (PBL Netherlands Environmental Assessment Agency, 2015).
 74. Boden, T. A., Marland, G. & Andres, R. J. *Global, Regional, and National Fossil-Fuel CO₂ Emissions* http://dx.doi.org/10.3334/CDIAC/00001_V2013 (Carbon Dioxide Information Analysis Center, 2013).
 75. Le Quéré, C. et al. Global Carbon Budget 2015. *Earth Syst. Sci. Data* **7**, 349–396 (2015).
- A yearly authoritative update of global anthropogenic carbon emissions, and natural and anthropogenic sinks.**
76. Armstrong, J. S. & Green, K. C. *Forecasting dictionary* <http://www.forecastingprinciples.com/index.php/forecasting-dictionary> [accessed 13 July 2012] (2012).
 77. Nakićenović, N. & Swart, R. *IPCC Special Report on Emissions Scenarios* <http://www.ipcc.ch/ipccreports/sres/emission/index.php?idp=0> (Cambridge Univ. Press, 2000).
 78. IEA. *World Energy Outlook 2014* <http://www.worldenergyoutlook.org/weo2014/> (International Energy Agency, 2014).
 79. Meinshausen, M. et al. National post-2020 greenhouse gas targets and diversity-aware leadership. *Nat. Clim. Change* **5**, 1098–1106 (2015).
 80. UNFCCC. *Review of the Implementation of Commitments and Other Provisions of the Convention*. Report No. FCCC/CP/2002/8, <http://unfccc.int/resource/docs/cop8/08.pdf>, 7 (UNFCCC, 2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all involved in the UNEP Emissions Gap Report, in particular J. McGlade, J. Alcamo and B. Metz, the members of the steering committee, all its authors, and the secretariat at UNEP-DTU, in particular A. Olhoff and J. Christensen. We acknowledge and are grateful for the data contributions of, and discussion with, all the global and national modelling groups, in particular M. Rocha, B. Hare, M. Schaeffer (Climate Analytics, Germany, Climate Action Tracker), L. Jeffery (PIK, Germany, Climate Action Tracker), P. van Breevoort (Ecofys, The Netherlands, Climate Action Tracker), A. Admiraal, M. Roelfsema, H. van Soest (PBL, The Netherlands), N. Forsell (IIASA, Austria), L. Cozzi, F. Kesicke (IEA, France), J. Cranston Turner, R. Boyd (LSE, UK), S. Dockweiler (DEA, Denmark), L. Siegel, E. Sawin (Climate Interactive, USA), A. Kitous, K. Keramidas, G. Grassi (JRC, European Commission), G. Iyer (PNL, USA), and the UNFCCC Secretariat. Without their contributions, this Perspective would have had a much more limited data basis. We thank S. Sterl (NewClimate Institute, Germany) and M. Roelfsema (PBL, The Netherlands) for assistance with data analysis, M. Strubegger (IIASA, Austria) for providing load factor estimates, V. Krey and O. Fricko (IIASA, Austria) for providing capture efficiencies for negative emissions technologies, and J. Cook for critical feedback on the manuscript and figures. We acknowledge and thank the International Institute for Applied Systems Analysis (IIASA) for hosting and maintaining the IPCC AR5 Scenario Database. J.R., N.H. and K.R. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 642147 (CD-LINKS). M.M. is the recipient of an Australian Research Council (ARC) Future Fellowship (grant number FT130100809).

Author Contributions The paper was initiated during discussions between M.d.E., N.H., H.W., J.R. and other members of the UNEP Gap Emissions Gap Report author team; all authors were involved in designing the research; the quantitative INDC analysis was coordinated by M.d.E., N.H. and H.F. with substantial contributions from T.F., H.W., R.S., F.S. and M.M.; the analysis of post-2030 implications was performed by J.R., with substantial contributions from K.R.; J.R. created all figures and was responsible for the overall coordination of the paper. J.R. and N.H. led the writing of the paper, with substantial contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.d.E. (Michel.denElzen@pbl.nl) or N.H. (n.hoehne@newclimate.org).

The Asian monsoon over the past 640,000 years and ice age terminations

Hai Cheng^{1,2}, R. Lawrence Edwards², Ashish Sinha³, Christoph Spötl⁴, Liang Yi⁵, Shitao Chen⁶, Megan Kelly², Gayatri Kathayat¹, Xianfeng Wang⁷, Xianglei Li¹, Xinggong Kong⁶, Yongjin Wang⁶, Youfeng Ning¹ & Haiwei Zhang¹

Oxygen isotope records from Chinese caves characterize changes in both the Asian monsoon and global climate. Here, using our new speleothem data, we extend the Chinese record to cover the full uranium/thorium dating range, that is, the past 640,000 years. The record's length and temporal precision allow us to test the idea that insolation changes caused by the Earth's precession drove the terminations of each of the last seven ice ages as well as the millennia-long intervals of reduced monsoon rainfall associated with each of the terminations. On the basis of our record's timing, the terminations are separated by four or five precession cycles, supporting the idea that the '100,000-year' ice age cycle is an average of discrete numbers of precession cycles. Furthermore, the suborbital component of monsoon rainfall variability exhibits power in both the precession and obliquity bands, and is nearly in anti-phase with summer boreal insolation. These observations indicate that insolation, in part, sets the pace of the occurrence of millennial-scale events, including those associated with terminations and 'unfinished terminations'.

The seasonal cycle of solar heating over Asia gives rise to the Asian monsoon (AM), a vast system of overturning atmospheric circulation that transports heat and moisture during boreal summer across the Indian Ocean and the tropical western Pacific into the Indian subcontinent and southeastern Asia, and as far as northern China and Japan¹ (Extended Data Fig. 1). Cave climate records have been important in characterizing AM changes and their causes. Such records demonstrate large and, in many cases, abrupt changes in monsoon intensity, inferred to have affected large swaths of Asia^{1–4}. A hallmark of these records is the precision with which age can be determined with modern U–Th dating methods⁵, thus allowing direct comparison with the orbital cycles without requiring orbital tuning^{1–4}. This approach, however, has been hindered by limited temporal coverage. Here we report a record from China, which, together with previously published data, covers the complete U–Th dating range from 640,000 years (640 kyr) ago to the present. Previous Chinese cave studies have demonstrated a close correspondence between changes in the AM and shifts in Northern Hemisphere summer insolation (NHSI) on orbital timescales^{3,6} and a close relationship between the AM and climate in the North Atlantic region on millennial scales^{1–4,7,8}. The latter has been used to correlate monsoon records with records from the North Atlantic region^{1–4,8}. Of note are Heinrich stadials (HSs) or ice rafted debris (IRD) events of North Atlantic origin, some of which coincide with Weak Monsoon Intervals (WMIs) in China^{4,8}. This correlation has been used to transfer the cave chronology to the marine oxygen isotope record, a strategy that has been important in establishing the timing of ice age terminations⁴ and which we apply here.

Our new $\delta^{18}\text{O}$ data from Sanbao Cave, China, allow us to establish the timing of Terminations (T) V through to VII in addition to the previously determined timing of terminations back to T-IV (ref. 4). First, this allows us to test ideas about the 100-kyr pacing of late Pleistocene ice age cycles⁹ and the degree to which termination timing is consistent with obliquity and/or precession forcing^{10–16}. Second, after removing the component of the AM that correlates with insolation, we examine

the residual suborbital variation⁸ over the full record. We show that some aspects of millennial-scale variability relate to orbital geometry. Third, the full record now crosses the Mid-Brunhes Event (MBE), when the character of CO_2 and ice volume cycles changed^{17–20}. We assess the degree to which these changes affected the AM. Fourth, we estimate the timing and duration of maximal AM strength over the Marine Isotope Stage (MIS) 11, a period of time which can be used as an analogue to the Holocene and future climate because of similar orbital geometry²¹.

Samples and results

Sanbao Cave is on the northern slope of Mt Shennongjia in central China (110° 26' E, 31° 40' N, elevation 1,900 metres above sea level). Mean annual temperature is 8 °C, and mean annual precipitation is 1,950 mm, 80% of which occurs during the summer (June to August). Four new stalagmites were collected ~1,500 m from the cave entrance. Samples were dated by a recently improved ²³⁰Th dating technique⁵, yielding precise age control (for example, ± 1.5 kyr at the time of T-V) (Extended Data Figs 2, 3 and Supplementary Table 1). New $\delta^{18}\text{O}$ measurements have a temporal resolution of between 200 and 70 years (average of ~120 years) (Extended Data Fig. 4 and Supplementary Table 1). The replication test (Extended Data Figs 2, 4) and other lines of reasoning^{1–4} suggest that speleothem $\delta^{18}\text{O}$ variability results from changes in the $\delta^{18}\text{O}$ of precipitation.

The climate interpretation of changes in the cave $\delta^{18}\text{O}$ records from China remains a subject of intense debate¹. However, most studies support one or both of the ideas presented in the original studies. Yuan *et al.*²² invoked Rayleigh fractionation to show that changes in the fraction of water vapour rained out between tropical sources and the cave site could account for the observed variability in the cave records. Most modelling studies (Liu *et al.*²³ and references therein) support this idea, although most refer to the process as 'upstream depletion'. Cheng *et al.*⁴ proposed that changes in the fraction of low $\delta^{18}\text{O}$ monsoon rainfall in annual totals could also explain the record. Recent theoretical²⁴ and empirical²⁵ studies support this idea, with the latter showing that both

¹Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an 710049, China. ²Department of Earth Sciences, University of Minnesota, Minneapolis, Minnesota 55455, USA.

³Department of Earth Sciences, California State University, Dominguez Hills, Carson, California 90747, USA. ⁴Institut für Geologie, Universität Innsbruck, A-6020 Innsbruck, Austria. ⁵State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China. ⁶College of Geography Science, Nanjing Normal University, Nanjing 210023, China. ⁷Earth Observatory of Singapore, Nanyang Technological University, 639798 Singapore.

processes can affect Chinese cave $\delta^{18}\text{O}$. For both, lower $\delta^{18}\text{O}$ implies higher spatially integrated monsoon rainfall between the tropical monsoon sources and the cave site and/or higher summer monsoon rainfall in the cave region. Thus, in this study, we use the terms ‘strong monsoon’ and ‘weak monsoon’ to refer to low and high cave $\delta^{18}\text{O}$, respectively, consistent with results from theoretical and empirical studies.

Our new records span from 640 to 330 kyr BP (before present, where present = AD 1950), which together with previous records (from 384 kyr BP to present), allow the construction of a composite AM $\delta^{18}\text{O}$ record, covering the past 640 kyr (Extended Data Fig. 4). The record is characterized by millennial-scale variations superimposed on a quasi-sine-wave-like orbital-scale variability that broadly tracks 21 July NHSI^{3,6} (Fig. 1). Removal of orbital-scale variations yields a record of the suborbital variability of the AM (the $\Delta\delta^{18}\text{O}$ record)⁸ (Extended Data Figs 5, 6). Detrending methods (for example, choice of insolation curve) could introduce artefacts in the $\Delta\delta^{18}\text{O}$ record, for which we tested by removing the orbital component of the record using insolation curves from a range of times encompassing the boreal summer months. Similar $\Delta\delta^{18}\text{O}$ power spectra independent of detrending curve suggest that this artefact is not significant. Detrending methods and sensitivity tests are described in the legends of Extended Data Figs 5, 6 and in Methods.

Timing and character of terminations

The gradual build-up and rapid termination of ice ages with an ~ 100 -kyr cycle are a well-known feature of the past ~ 650 kyr (ref. 9). Although both glacial cycles and changes in eccentricity share common spectral power, the latter generates negligible change in insolation, thus presenting an enduring climate puzzle—the so-called ‘100-kyr problem’. A number

of hypotheses have been put forth to address this problem. One hypothesis explains the 100-kyr cycle as an average of 4 to 5 discrete precession cycles, with missed beats in between^{10,11}. Another invokes 2 to 3 obliquity cycles^{12,13}, again with missed beats. Yet another invokes a combination of both obliquity and precession^{14–16}. Others call for interactions involving internal oscillations in the Earth system^{26,27}.

Cheng *et al.*⁴ have shown that each of the last four terminations is characterized by one or two WMIs, which coincide with HSs observed in North Atlantic marine cores^{7,8}. Abrupt WMI endings are synchronous with abrupt increases in atmospheric CH_4 in Antarctic ice cores⁴. Using these cave–marine and cave–ice core correlations, Cheng *et al.*⁴ placed the events observed in marine and ice cores on a cave chronology and made the following observations: the WMIs correlated with a good portion of each marine termination; the WMIs and the marine terminations took place at a time of rising NHSI; and most of the CO_2 rise associated with each termination took place during the WMIs (Extended Data Fig. 7). On the basis of these observations, Cheng *et al.*⁴ suggested that for each termination, the rise in insolation triggered the initial melting of the ice sheets. The North Atlantic cold anomaly that resulted from input of ice and meltwater rearranged oceanic and atmospheric circulation, causing the WMIs and resulting in the rise in atmospheric CO_2 . The latter, along with a continuing rise in insolation drove the termination^{4,28}.

The unparalleled length and temporal precision of our cave record allow us to extend the aforementioned approach to robustly test ideas about the classic ‘100-kyr problem’. Our data indicate that glacial terminations T-VII to T-V were also associated with WMIs (Fig. 2). The T-V WMI occurred between $\sim 430.5 \pm 1.5$ and $\sim 426 \pm 2$ kyr BP. The T-VI WMI, centred at 532.3 ± 3.5 kyr BP, has a duration of ~ 4.5 kyr,

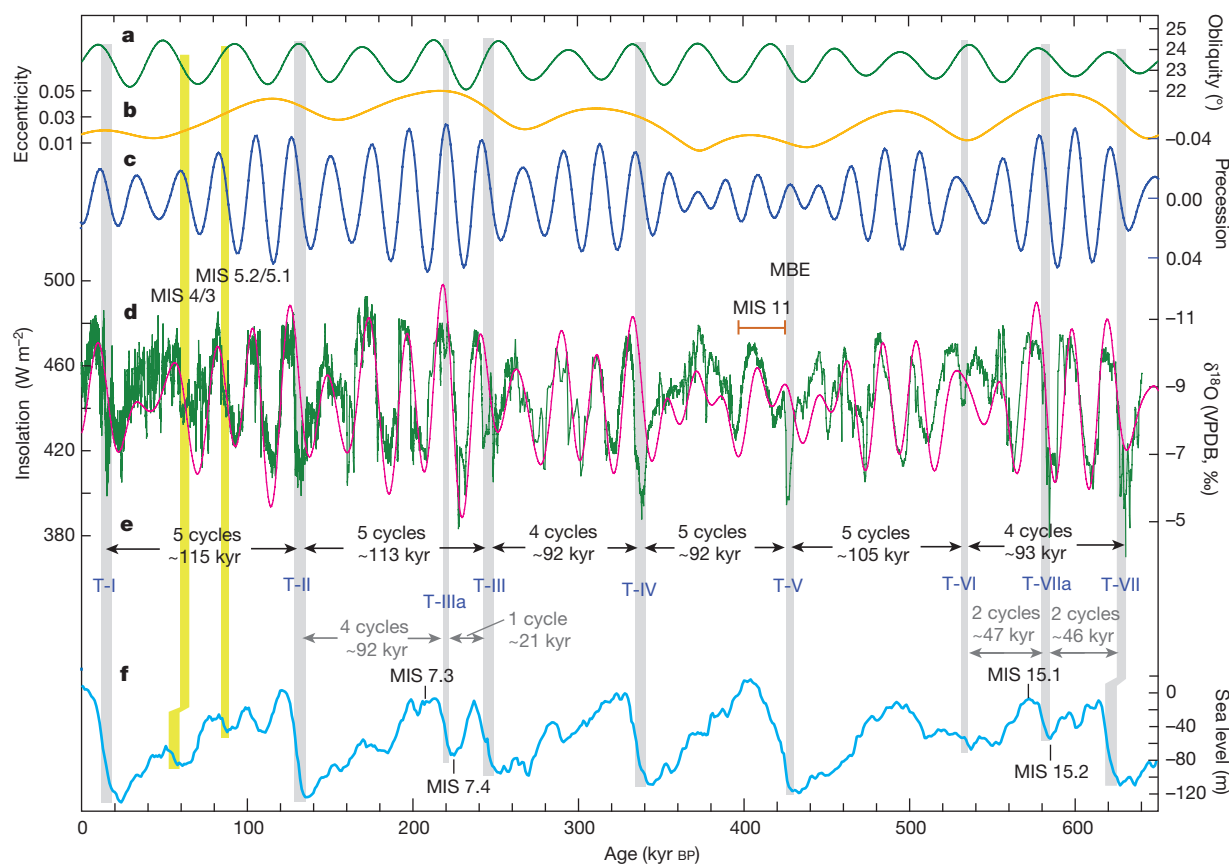


Figure 1 | Asian monsoon variations in the context of the Earth's orbital parameters. a–c, Changes in obliquity (a), eccentricity (b) and precession⁴⁵ (c). d, The composite AM $\delta^{18}\text{O}$ record (green; this study) and 21 July insolation at 65°N (ref. 45; pink). e, Termination pacing and duration. Vertical bars mark the timing of WMIs correlated to glacial terminations

(grey) and two similar events (MIS 4/3 and 5.2/5.1 transitions) (yellow). The timing of T-IIIa-WMI in this study differs from the one described in ref. 4, although we consider the latter a plausible alternative (see main text and Extended Data Fig. 9). f, The composite sea level¹⁷. The timings of MBE, MIS 11, 7.3, 7.4, 15.1 and 15.2 are also depicted.

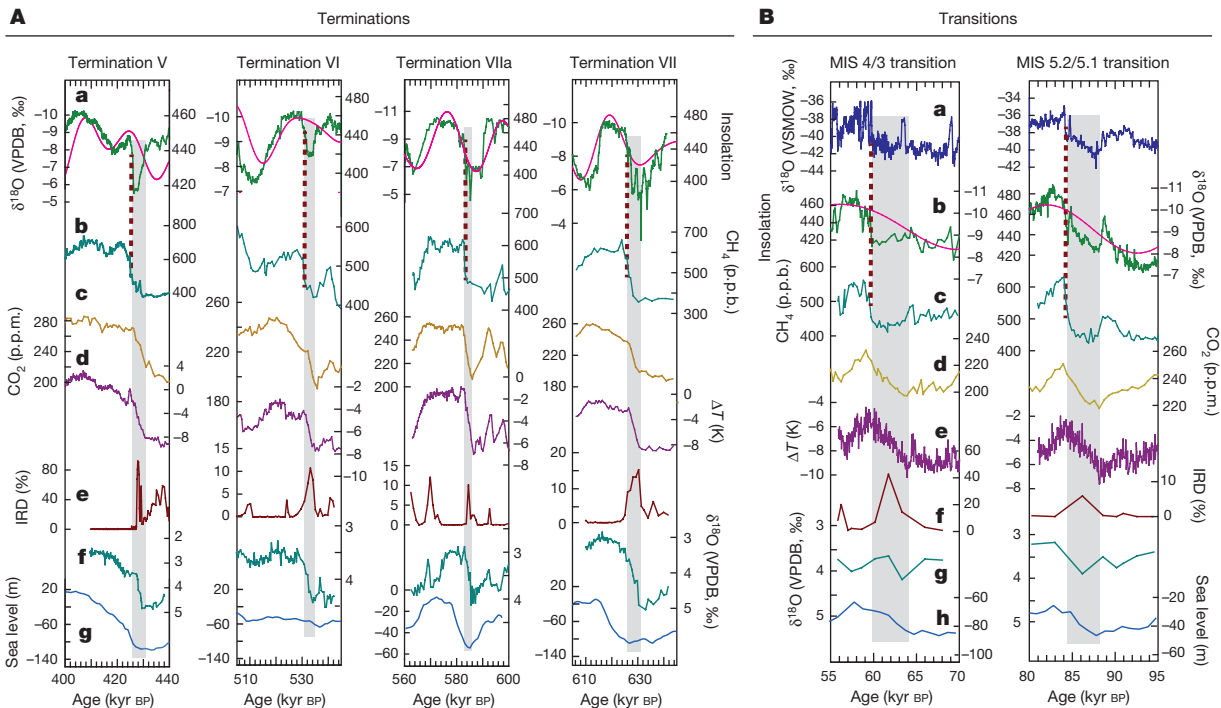


Figure 2 | Comparison of climate events surrounding terminations and other two millennial-scale events. **A**, Termination events surrounding T-V to T-VII. **a**, 21 July insolation at 65°N (pink, W m^{-2})⁴⁵ and the AM $\delta^{18}\text{O}$ record (green). **b**, **d**, EDC CH_4 (ref. 19) and relative temperature¹⁸ records, respectively. **c**, **e**, The composite CO_2 record²⁰. The ice core ages (EDC3 chronology⁴⁶) around T-VI and T-VIIa are shifted to the older side by 3 and 2 kyr, respectively, to match the abrupt AM and CH_4 changes. **e**, **f**, Marine ODP980⁴⁷ (T-V) and U1314⁴⁸ (T-VI to T-VII) IRD and benthic $\delta^{18}\text{O}$ records, respectively. The marine records around T-V, T-VI, T-VIIa and T-VII are shifted to the older side by 6 kyr, the younger side by 2 kyr, older side by 2 kyr and older side by 3 kyr, respectively, to match corresponding IRD events to WMIs. **g**, The composite sea level¹⁷. **B**, Climate events surrounding MIS 4/3 and 5.2/5.1 transitions.

assuming a linear growth rate of sample SB-32 around this time period. The T-VII WMI is broadly similar to that of T-II, T-IV and T-V, ending abruptly at 627 ± 6 kyr BP. Each coincides presumably with a major HS, a sea level rise marking each glacial termination, a CO_2 rise, and an Antarctic temperature rise (Fig. 2). These observations, now for the past seven terminations, support the hypothesis that rising NHSI triggers an initial ice-sheet disintegration, which in turn perturbs the oceanic and atmospheric heat and carbon cycles, resulting in a CO_2 increase, which further drives the termination^{4,28}.

All seven terminations occurred during the rising limbs of NHSI separated by 4, 5, 5, 4, 5, and 5 precession cycles. The durations between successive terminations from T-VII to T-I were about 93, 105, 92, 92, 113 and 115 kyr, respectively, rather than strict 100-kyr cycles (Fig. 1). Thus, the '100-kyr cycle' is an approximate average of intervals that are generally a little longer or shorter than 100 kyr. In addition, we characterize two 'extra terminations' revealed in marine records as follows: T-IIIa (ref. 12) occurred one precession cycle after T-III (between MIS 7.4 and 7.3), and T-VIIa occurred two precession cycles after T-VII (between MIS 15.2 and 15.1) (Fig. 1). Both exhibit a pattern of events similar to the seven main terminations, and large and comparable marine $\delta^{18}\text{O}$ or sea level changes¹⁷ (Figs 1, 2).

Each of the nine terminations is separated from adjacent terminations by an integral number of precession cycles. Thus, insolation intensity is critical in controlling the timing of terminations. In contrast, termination timing does not exhibit an obvious relationship to any particular portion of the obliquity or eccentricity cycles. For example, terminations take place when eccentricity is high (>0.02 : T-II, T-IIIa, T-III, T-IV, T-VIIa, T-VII) and when it is low (<0.02 : T-I, T-V,

a, Greenland ice core (NGRIP) $\delta^{18}\text{O}$ record⁴⁹. **b**, 21 July insolation at 65°N (pink, W m^{-2})⁴⁵ and AM $\delta^{18}\text{O}$ record (green). **c**, **e**, EDC CH_4 (ref. 19) and relative temperature¹⁸ records, respectively. **d**, The composite CO_2 record²⁰. The ice core ages around MIS 4/3 and 5.2/5.1 transitions are shifted to the older side by 0.5 and 1 kyr respectively to match the abrupt AM and CH_4 changes. **f**, **g**, ODP980 IRD and benthic $\delta^{18}\text{O}$ records⁷, respectively. On the basis of correlations between IRD events and WMIs, the ODP980 ages are shifted to the younger side by ~ 4 and 3 kyr around MIS 4/3 and 5.2/5.1 transitions, respectively. **h**, The composite sea level¹⁷. Its chronology around the MIS 4/3 transition is shifted to the older side by ~ 5 kyr. Vertical grey bars indicate WMIs and corresponding events (CO_2 , Antarctic temperature, and North Atlantic IRD events). Dashed lines depict correlations between abrupt intensification in AM and CH_4 jump.

T-VI) (Fig. 1). Thus, changes in eccentricity are not, in a direct fashion, responsible for the pacing of the '100-kyr cycle'. Similarly, terminations take place as obliquity is increasing (T-IIIa, T-V, T-VII), decreasing (T-III, T-VI), and at peak or near peak values (T-I, T-II, T-IV, T-VIIa). Thus, as with eccentricity, obliquity does not pace terminations in a precise fashion. However, as none of the nine events takes place when obliquity is substantially below average, obliquity may play some causal role^{12–16}.

Pacemaker of millennial-scale events

A number of studies have focused on the cause of skipped precession and/or obliquity beats^{10,12,15} between terminations. More recent studies have referred to the events at some of these 'missed beats' as 'low-amplitude versions of terminations'²⁹, 'failed terminations'⁸, or 'unfinished terminations'³⁰. For example, the MIS 4/3 and MIS 5.2/5.1 transitions took place at times of NHSI rise and have many of the features of full terminations, including periods of unusually weak monsoon (Fig. 2)—although, akin to T-VI, they do not result in extended interglacial conditions, but are characterized by significant sea level rise with larger marine $\delta^{18}\text{O}$ shifts than T-VI¹⁷ (Figs 1, 2). Explanations for full versus small-scale terminations focus on interplays and feedbacks among factors internal to the climate system, such as ice volume dynamics, in addition to orbital forcing^{15,31–34}.

As all full terminations and, at least, some small-scale terminations are associated with millennial-scale intervals of unusually weak monsoon, we examine the suborbital component of the full record to probe its relationship to orbital geometry. We start by removing orbital-scale frequencies to create a residual $\Delta\delta^{18}\text{O}$ record (Fig. 3), in which

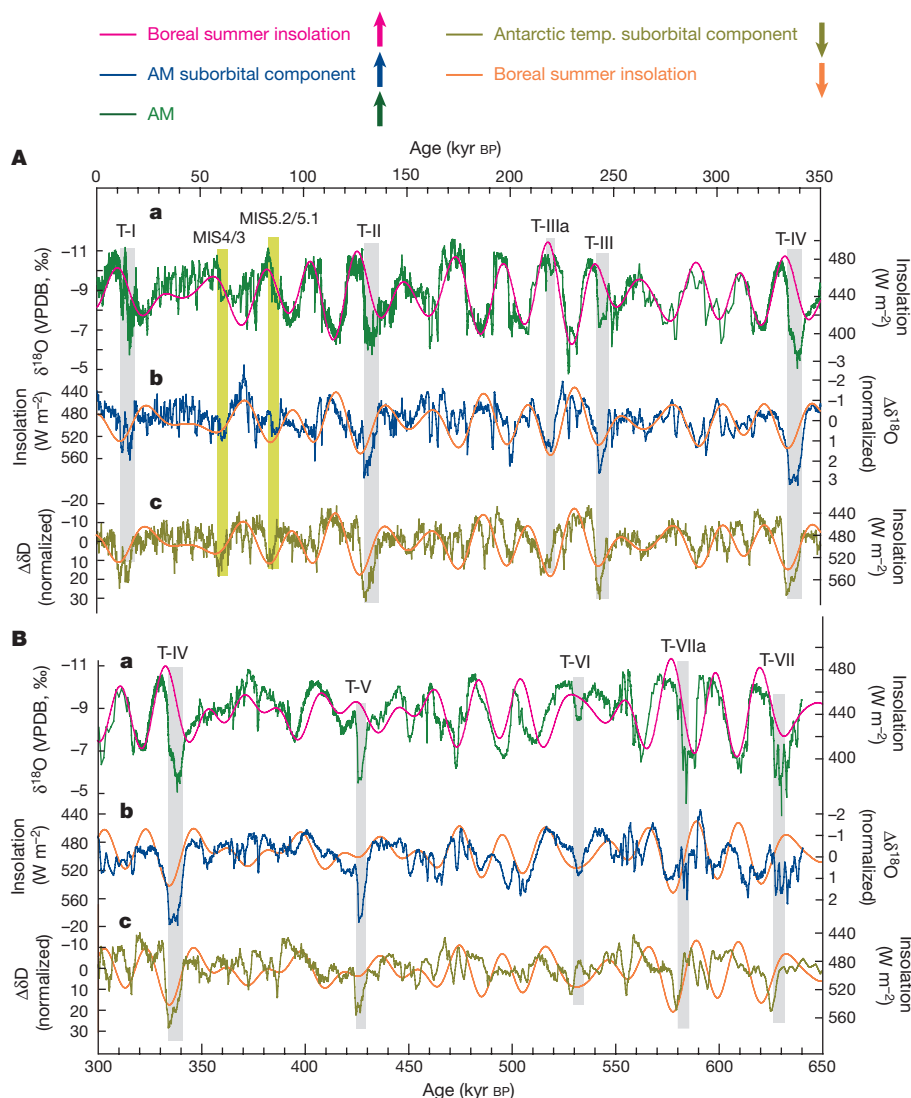


Figure 3 | Comparison of suborbital AM and Antarctic temperature variations. **A**, Interval from 350 to 0 kyr BP. **B**, Interval from 650 to 300 kyr BP. In both **A** and **B**: **a**, the composite AM $\delta^{18}\text{O}$ record (green) and 21 July insolation at 65°N (pink)⁴⁵; **b**, suborbital AM variation ($\Delta\delta^{18}\text{O}$, detrended from the composite $\delta^{18}\text{O}$ record by subtracting 21 July insolation at 65°N ; see Methods); **c**, Antarctic suborbital temperature

variation ($\Delta\delta\text{D}$ record detrended from the δD record¹⁸; see Methods). Orange curves show 21 June insolation at 65°N (ref. 45) on a reversed scale (increasing down) for comparison. A remarkable similarity is evident between suborbital variations in AM $\Delta\delta^{18}\text{O}$ and Antarctic $\Delta\delta\text{D}$ records (also see Extended Data Fig. 8).

suborbital variability is clearly observed⁸. Prominent features are the aforementioned WMIs, millennial-scale features characterized by large positive anomalies in $\delta^{18}\text{O}$ at terminations. A closer look shows many smaller amplitude, millennial-scale, high- $\delta^{18}\text{O}$ events throughout, including those correlated with small-scale terminations. We infer that these events may share a similar origin with the WMIs, since the pattern of events surrounding these smaller events is indeed similar to terminations, including some for which the marine $\delta^{18}\text{O}$ /sea level shift is comparable or larger than that of T-VI (Fig. 2). Specifically, we infer that the smaller amplitude events are caused by the decay of the northern ice sheets, resulting in the flux of ice and meltwater into the North Atlantic; the ensuing slowdown in Atlantic Meridional Overturning Circulation (AMOC) generates a cold anomaly over the North Atlantic. Through an atmospheric teleconnection, the cold anomaly results in a weaker AM, recorded as a high $\delta^{18}\text{O}$ anomaly in our record (Fig. 2)^{1,4,28}.

Spectral analysis of the $\Delta\delta^{18}\text{O}$ record reveals strong power in the precession band, with the suborbital component of the AM close to anti-phased with (that is, $\sim 180^\circ$ from) 21 June insolation (Fig. 4b). The suborbital spectrum also exhibits weaker but significant power in the obliquity band, again with the suborbital component close to

anti-phased with obliquity. Thus, insolation modulates the suborbital component of AM variability, but in the opposite sense from the more direct control of orbital-scale variability of the AM. We take these observations to indicate that high NHSI, whether in the precession or obliquity bands, favours the disintegration of the northern ice sheets and the release of ice and meltwater into the North Atlantic. This signal then propagates through the ocean and atmosphere as outlined above (and previously described for WMIs), resulting in the weakening of the monsoon^{4,28}. Inspection of Fig. 3 shows that the relationship is pervasive throughout the record.

We have performed a similar analysis on the Antarctic δD (a temperature proxy) record¹⁸, by first removing low frequency, orbital-scale variability, then examining the power spectrum of the resulting suborbital variability³⁵ (Fig. 4 and Extended Data Fig. 8). The suborbital-scale Antarctic temperature and AM records are remarkably similar, with the suborbital components of the AM and Antarctic temperature anti-correlated. We attribute this relationship to the bi-polar seesaw mechanism³⁶. As input of meltwater and ice into the North Atlantic is the presumed source of both sets of suborbital variability, it is not surprising that the suborbital component of Antarctic temperature

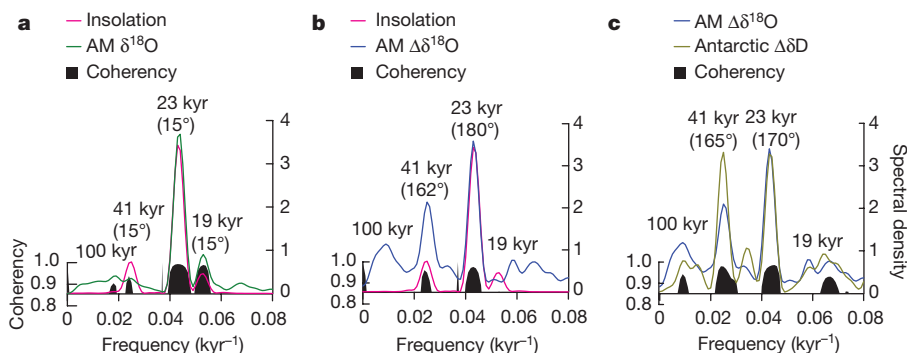


Figure 4 | Cross-spectral comparison. Compared are insolation, AM, detrended (suborbital) AM and detrended (suborbital) Antarctic δD records over the past 640 kyr BP. **a**, Spectral analysis results of 21 July insolation at 65° N (ref. 45; pink) and AM $\delta^{18}O$ (green), and their coherence spectra (black). **b**, Spectral analysis results of 21 June insolation at 65° N (ref. 45; pink) and detrended AM $\Delta\delta^{18}O$ (blue), and their coherence spectra (black). **c**, Spectral analysis results of detrended AM $\Delta\delta^{18}O$ (blue) and detrended Antarctic $\Delta\delta D$ records (olive), and their coherence spectra (black). Coherence spectra are at 80% confidence level.

Numbers in parentheses show phase differences in degrees between two spectra at precession (~ 23 and 19 kyr) and obliquity (~ 41 kyr) bands. Precession periodicity of ~ 23 kyr is significant in all four datasets. The variations in both AM $\delta^{18}O$ and Antarctic $\Delta\delta D$ records also show significant power at obliquity band. Cross-spectrum analyses reveal a phase coherence at precession band between AM $\delta^{18}O$ record and 21 July insolation and between Antarctic $\Delta\delta D$ and 21 June insolation. AM $\Delta\delta^{18}O$ is nearly anti-phased with Antarctic $\Delta\delta D$ and 21 June insolation at precession band.

variability also exhibits significant power in the precession and obliquity bands nearly in phase with NHSI. These results suggest that the change of NHSI induced by precession (with lesser involvement of obliquity) is the major external pacemaker of both conventional and the 'small-scale' termination events, which coincide with the large-scale weakening of the AM and warming in Antarctica^{4,8,28,30}.

While the sequence of events surrounding each positive AM $\delta^{18}O$ anomaly is similar, the relative amplitudes of these events vary (Fig. 3). Hence, the key issue is not why terminations are spaced by some discrete number of precession cycles, but rather why some of the events developed ultimately into full terminations. The possibility of a certain event evolving into a full termination may depend on the state of the climate system at that time—for example, favourable internal interplays/feedbacks of ice-sheet dynamics and size, insolation, and ocean

and CO_2 feedbacks^{4,28,30–34}. For example, as pointed out by Raymo³¹, a large and isostatically-compensated ice sheet would be relatively vulnerable to decay. In this scenario, precession cycles would be skipped or a 'termination' would be unfinished until the ice sheets were in such a state, on average after about 100 kyr. Regardless of exact mechanisms, the WMIs and smaller analogues to WMIs are indeed paced primarily by precession cycles, suggesting that the 100 kyr in the '100-kyr problem' represents an approximate mean of integral numbers of precession cycles between terminations.

Mid-Brunhes event and MIS 11

Our record covers much of the time since the so-called Mid-Pleistocene revolution¹¹ and crosses the MBE³⁷. The MBE is contemporaneous with T-V, a portion of which correlates with a WMI in our record. The

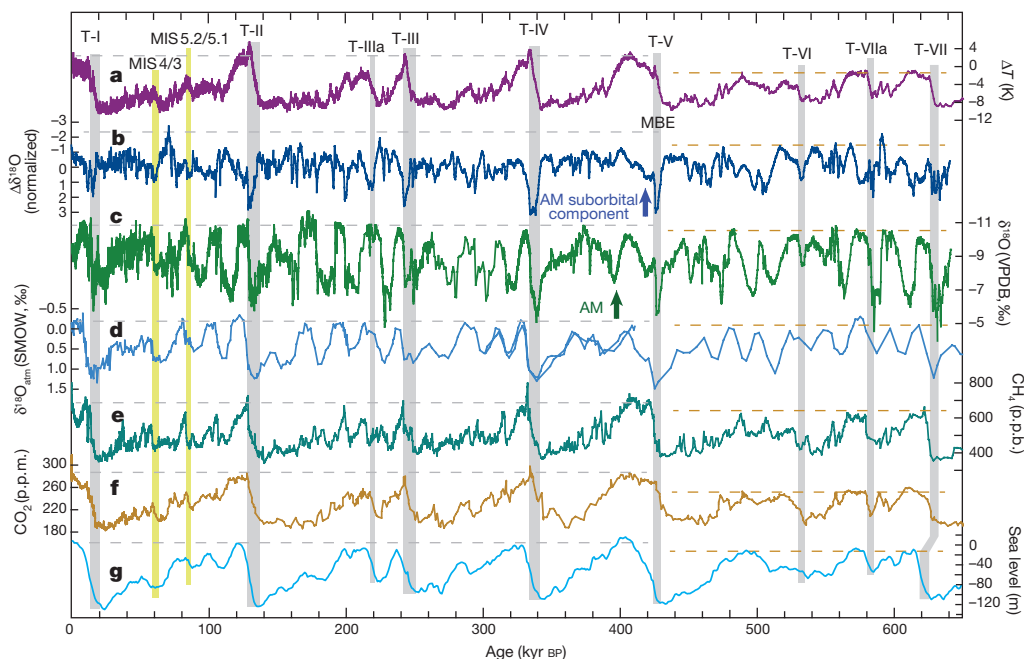


Figure 5 | Comparison of climate records across the MBE.

a, d, e, Relative temperature¹⁸, $\delta^{18}O_{atm}$ (ref. 50) and CH_4 (ref. 19) records, respectively, from Antarctic EDC ice core records (EDC3 chronology⁴⁶). **b**, The suborbital variations in the AM $\Delta\delta^{18}O$ record. **c**, The orbital variations in the AM $\delta^{18}O$ record. **f**, The composite CO_2 record²⁰.

g, The composite sea level¹⁷. Horizontal dashed lines indicate interglacial amplitude shift across the MBE. Vertical bars indicate WMIs and associated T-I to T-VII (grey) as well as MIS 4/3 and 5.2/5.1 transitions (yellow).

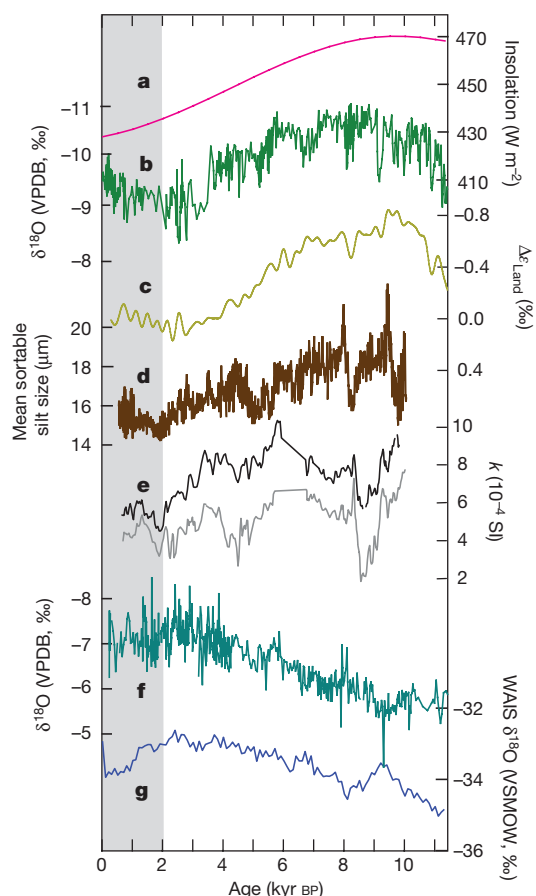


Figure 6 | Comparison among Holocene records. **a**, 21 July insolation at 65° N (ref. 45). **b**, The composite AM $\delta^{18}\text{O}$ records. **c**, The Holocene terrestrial $^{18}\text{O}/^{16}\text{O}$ fractionation variation³⁹. **d**, The mean grain size of sortable silt (10–63 μm) for North Atlantic sediment cores MD2251 and MD2024, a proxy of North Atlantic Deep Water formation⁴³. **e**, North Atlantic deep-flow dynamic proxy (low field magnetic susceptibility, k) records from cores CH77-02 (black) and MD08-3182Cq (grey)⁴⁴. **f**, The South American monsoon record⁴⁰. **g**, The $\delta^{18}\text{O}$ record (WAIS, a temperature proxy) from western Antarctica⁴¹. Vertical bar depicts the '2-kyr shift'.

amplitude of glacial–interglacial cycles in CO_2 , Antarctic temperature, and global ice volume (or sea level) increased substantially after the MBE. In contrast, interglacial CH_4 , $\delta^{18}\text{O}_{\text{atm}}$, and AM intensity in our records are only slightly enhanced across the MBE (Fig. 5), similar to another hydrological record reported previously from low latitude³⁸. Thus, changes in the AM behaviour across the MBE were not as large as changes at high latitudes and in CO_2 . In addition, there are numerous examples of high AM intensity during glacial periods when ice volume was large and CO_2 was low (Fig. 5). Furthermore, the initial increase of the AM tends to start at each insolation minimum and in some cases does not appear to correlate with any obvious ice-volume and/or CO_2 changes (Extended Data Fig. 9). These observations indicate that, at the orbital scale, the AM and associated climate and atmospheric chemistry change are largely controlled by NHSI⁶. However, the post-MBE WMIs associated with glacial terminations are much more intense (Fig. 5b). It is possible that this change results from a shift in the response time of the AM to NHSI. However, in our view, a more likely cause is the observed shift to higher maximum ice volume resulting in higher ice and meltwater flux during terminations and larger perturbations of the climate system, including more extreme WMIs. In this light, T-V has the first of the large post-MBE-style WMIs.

Our new data also include a well-dated record of AM variability during MIS 11. Orbital parameters for MIS 11 are closer to those of

the Holocene than for other recent interglacial periods, making it an analogue for the Holocene and future climate²¹. However, uncertainties remain regarding its timing and duration. We estimate the beginning of AM MIS 11 to be the abrupt end of the T-V WMI at 426 ± 1 kyr BP. If we take the end to be the AM minimum at 396 ± 3 kyr BP (about the time of the half-height of the benthic $\delta^{18}\text{O}$ MIS 11/10 shift¹⁷), the duration is 30 ± 4 kyr. If we take the end to be the time of the half-height of the AM shift after the MIS 11 AM peak (at 399 ± 3 kyr BP), the duration is 27 ± 4 kyr. These estimates are consistent within age uncertainties with the duration of MIS 11 defined by other less precisely dated records²¹.

The '2-kyr shift'

Over the past ~ 2 kyr the AM has increased in an anomalous fashion relative to the downward trend in NHSI (Fig. 6 and Extended Data Fig. 10). This trend relates to climate change elsewhere around the world in much the same way as the millennial-scale shifts over the past 640 kyr. We refer to this late Holocene anomaly as the '2-kyr shift'. The '2-kyr shift' is more than just a regional shift in the AM as evidenced by an in phase, positively-correlated shift in the $\delta^{18}\text{O}$ of atmospheric O_2 (ref. 39), which integrates a broad swath of the globe and may, in large part, respond to AM intensity. The AM shift also anti-correlates with records of the South American monsoon⁴⁰ and temperature over some parts of Antarctica⁴¹ (Fig. 6) — a pattern similar to the one observed for millennial-scale events throughout much of the past several hundred kyr. A change in AMOC has been suggested as a key process that may explain aspects of such millennial-scale changes⁴². In this regard, two studies^{43,44} show generally decreasing AMOC for several millennia before ~ 2 kyr BP, then constant or perhaps slightly increasing AMOC for the last 2 kyr (Fig. 6). Thus, it is plausible that the origin of the '2-kyr shift' is a progressive increase in the rate of AMOC over the past 2 kyr. Nevertheless, observational and modelling studies are critically needed to further assess the '2-kyr shift'.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 January; accepted 18 May 2016.

- Cheng, H., Sinha, A., Wang, X. F. & Cruz, F. W. The global paleomonsoon as seen through speleothem records from Asia and South America. *Clim. Dyn.* **39**, 1045–1062 (2012).
- Wang, Y. J. et al. A high-resolution absolute-dated Late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
- Wang, Y. J. et al. Millennial- and orbital-scale changes in the East Asian Monsoon over the past 224,000 years. *Nature* **451**, 1090–1093 (2008).
- Cheng, H. et al. Ice age terminations. *Science* **326**, 248–252 (2009).
- Cheng, H. et al. Improvements in ^{230}Th dating, ^{230}Th and ^{234}U half-life values, and U-Th isotopic measurements by multi-collector inductively coupled plasma mass spectrometry. *Earth Planet. Sci. Lett.* **371–372**, 82–91 (2013).
- Kutzbach, J. E., Liu, X. D., Liu, Z. Y. & Chen, G. S. Simulation of the evolutionary response of global summer monsoons to orbital forcing over the past 280,000 years. *Clim. Dyn.* **30**, 567–579 (2008).
- McManus, J. F., Oppo, D. W. & Cullen, J. L. A 0.5-million-year record of millennial-scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
- Barker, S. et al. 800,000 years of abrupt climate variability. *Science* **334**, 347–351 (2011).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Raymo, M. E. Glacial puzzles. *Science* **281**, 1467–1468 (1998).
- Maslin, M. A. & Ridgwell, A. in *Early-Middle Pleistocene Transitions: The Land-Ocean Evidence* (eds Head, M. J. & Gibbard, P. L.) 19–34 (Vol. 247, Spec. Publ. Geol. Soc. Lond., 2005).
- Huybers, P. & Wunsch, C. Obliquity pacing of the late Pleistocene glacial terminations. *Nature* **434**, 491–494 (2005).
- Bintanja, R. & van de Wal, R. S. W. North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. *Nature* **454**, 869–872 (2008).
- Imbrie, J. & Imbrie, J. Modeling the climatic response to orbital variations. *Science* **207**, 943–953 (1980).
- Paillard, D. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature* **391**, 378–381 (1998).
- Huybers, P. Combined obliquity and precession pacing of late Pleistocene deglaciations. *Nature* **480**, 229–232 (2011).

17. Spratt, R. M. & Lisiecki, L. E. A Late Pleistocene sea level stack. *Clim. Past Discuss.* **11**, 3699–3728 (2015).
18. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
19. Loulergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH₄ over the past 800 000 years. *Nature* **453**, 383–386 (2008).
20. Bereiter, B. *et al.* Revision of the EPICA Dome C CO₂ record from 800 to 600 kyr before present. *Geophys. Res. Lett.* **42**, 542–549 (2015).
21. Candy, I., Schreve, D. C., Sherriff, J. & Tye, G. J. Marine Isotope Stage 11: Palaeoclimates, palaeoenvironments and its role as an analogue for the current interglacial. *Earth Sci. Rev.* **128**, 18–51 (2014).
22. Yuan, D. X. *et al.* Timing, duration, and transitions of the Last Interglacial Asian monsoon. *Science* **304**, 575–578 (2004).
23. Liu, Z. Y. *et al.* Chinese cave records and the East Asia Summer Monsoon. *Quat. Sci. Rev.* **83**, 115–128 (2014).
24. Chiang, J. C. H. *et al.* Role of seasonal transitions and westerly jets in East Asian paleoclimate. *Quat. Sci. Rev.* **108**, 111–129 (2015).
25. Orland, I. J. *et al.* Direct measurements of deglacial monsoon strength in a Chinese stalagmite. *Geology* **43**, 555–558 (2015).
26. Saltzman, B., Hansen, A. & Maasch, K. The late Quaternary glaciations as the response of a three-component feedback system to Earth-orbital forcing. *J. Atmos. Sci.* **41**, 3380–3389 (1984).
27. Toggweiler, J. R. Origin of the 100,000-year timescale in Antarctic temperatures and atmospheric CO₂. *Paleoceanography* **23**, PA2211 (2008).
28. Denton, G. H. *et al.* The last glacial termination. *Science* **328**, 1652–1656 (2010).
29. Wolff, E. W., Fischer, H. & Röthlisberger, R. Glacial terminations as southern warmings without northern control. *Nat. Geosci.* **2**, 206–209 (2009).
30. Schaefer, J. M. *et al.* The southern glacial maximum 65,000 years ago and its unfinished termination. *Quat. Sci. Rev.* **114**, 52–60 (2015).
31. Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12**, 577–585 (1997).
32. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles 2. The 100,000-year cycle. *Paleoceanography* **8**, 699–735 (1993).
33. Parrenin, F. & Paillard, D. Amplitude and phase of glacial cycles from a conceptual model. *Earth Planet. Sci. Lett.* **214**, 243–250 (2003).
34. Abe-Ouchi, A. *et al.* Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume. *Nature* **500**, 190–193 (2013).
35. Siddall, M., Rohling, E. J., Blunier, T. & Spahni, R. Patterns of millennial variability over the last 500 ka. *Clim. Past* **6**, 295–303 (2010).
36. Broecker, W. S. Paleoocean circulation during the last deglaciation: a bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
37. Jansen, J. H. F., Kuijpers, A. & Troelstra, S. R. A Mid-Brunhes climatic event: long-term changes in global atmosphere and ocean circulation. *Science* **232**, 619–622 (1986).
38. Meckler, A. N., Clarkson, M. O., Cobb, K. M., Sodemann, H. & Adkins, J. F. Interglacial hydroclimate in the tropical West Pacific through the late Pleistocene. *Science* **336**, 1301–1304 (2012).
39. Severinghaus, J. P., Beaudette, R., Headly, M. A., Taylor, K. & Brook, E. J. Oxygen-18 of O₂ records the impact of abrupt climate change on the terrestrial biosphere. *Science* **324**, 1431–1434 (2009).
40. van Breukelen, M. R., Vonhof, H. B., Hellstrom, J. C., Wester, W. C. G. & Kroon, D. Fossil dripwater in stalagmites reveals Holocene temperature and rainfall variation in Amazonia. *Earth Planet. Sci. Lett.* **275**, 54–60 (2008).
41. WAIS Divide Project Members. Onset of deglacial warming in West Antarctica driven by local orbital forcing. *Nature* **500**, 440–444 (2013).
42. Broecker, W. S. & Denton, G. H. The role of ocean-atmosphere reorganizations in glacial cycles. *Geochim. Cosmochim. Acta* **53**, 2465–2501 (1989).
43. Hoogakker, B. A. A. *et al.* Dynamics of North Atlantic Deep Water masses during the Holocene. *Paleoceanography* **26**, PA4214 (2011).
44. Kissel, C., Toer, A. V., Laj, C., Cortijo, E. & Michel, E. Variations in the strength of the North Atlantic bottom water during Holocene. *Earth Planet. Sci. Lett.* **369–370**, 248–259 (2013).
45. Berger, A. Long-term variations of caloric insolation resulting from the Earth's orbital elements. *Quat. Res.* **9**, 139–167 (1978).
46. Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).
47. Alonso-Garcia, M. *et al.* Ocean circulation, ice sheet growth and interhemispheric coupling of millennial climate variability during the Mid-Pleistocene (ca. 800–400 ka). *Quat. Sci. Rev.* **30**, 3234–3247 (2011).
48. Wright, A. K. & Flower, B. P. Surface and deep ocean circulation in the subpolar North Atlantic during the mid-Pleistocene revolution. *Paleoceanography* **17**, 1068 (2002).
49. Wolff, E. W., Chappellaz, J., Blunier, T., Rasmussen, S. O. & Svensson, A. Millennial-scale variability during the last glacial: The ice core record. *Quat. Sci. Rev.* **29**, 2828–2838 (2010).
50. Landais, A. *et al.* What drives the millennial and orbital variations of $\delta^{18}\text{O}_{\text{atm}}$? *Quat. Sci. Rev.* **29**, 235–246 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by China grants NBRP 2013CB955902, NSFC 41230524, 4157020432 and 41561144003, US NSF grants 0502535, 1103404, 0823554, 1003690, 1137693 and 1317693 and Singapore grant NRF-NRFF2011-08. We thank M. Siddall for help with analysis of the millennial-scale variability of the Antarctic temperature record and A. P. Roberts for converting the ice core chronology.

Author Contributions H.C. designed the research and experiments; H.C., R.L.E. and A.S. wrote the manuscript, which was edited by all of the co-authors; L.Y., S.C. and A.S. did the spectral analysis; X.K., Y.W. and S.C. provided the cave samples; H.C. did the ²³⁰Th dating work; and C.S., X.K., M.K., Y.N. and H.Z. contributed to oxygen isotope measurements. All authors discussed the results and provided input to the manuscript and technical aspects of the laboratory analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. (cheng021@mail.xjtu.edu.cn) or R.L.E. (edwar001@umn.edu).

Reviewer Information *Nature* thanks C. Buizert, A. N. Meckler and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

²³⁰Th dating. Four stalagmite samples, SB-12, SB-14, SB-32 and SB-58, were collected from Sanbao Cave, Hubei, China. The sampling location is ~1,500 m from the entrance, marked by a relative humidity of ~100%. The stalagmites were cut into halves along their growth axes and polished. A total of 196 sub-samples (25, 69, 84 and 18 for SB-12, SB-14, SB-32 and SB-58, respectively) were drilled for ²³⁰Th dating (Supplementary Table 1). In addition, 9 dates were also obtained for stalagmite D8 from Dongge Cave, China. The dating work was performed at the Minnesota Isotope Laboratory, University of Minnesota (Sanbao samples) and the Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an, China (Dongge samples). The ²³⁰Th dating techniques are essentially identical in the two laboratories. All measurements were made on Thermo-Finnigan Neptune multi-collector inductively coupled plasma mass spectrometers using the recently improved technique⁵. We use standard chemistry procedures to separate U and Th as described in ref. 51. The isotope dilution method with a triple-spike ²²⁹Th–²³³U–²³⁶U was employed to correct for instrumental fractionation and determine U and Th isotopic ratios and concentrations. The instrumentation, standardization and half-lives are reported in refs 5 and 52. All U and Th isotopes were measured either on the Faraday cups (larger sample size) or on a MasCom multiplier behind the retarding potential quadrupole in the peak-jumping mode (smaller sample size). We followed similar procedures of characterizing the multiplier as described in ref. 52. Uncertainties in U and Th isotopic data were calculated offline at the 2σ level, including corrections for chemistry/instrument blanks, multiplier dark noise, abundance sensitivity, tails, and contents of the same four nuclides in the spike solution^{5,52}. Corrected ²³⁰Th ages assume an initial ²³⁰Th/²³²Th atomic ratio of $(4.4 \pm 2.2) \times 10^{-6}$, the values for a material at secular equilibrium with the bulk Earth ²³²Th/²³⁸U value of 3.8. The correction is negligible because the samples used in this study have high U and low Th contents. The age model for each stalagmite is established by either linear interpolation or polynomial fitting (Extended Data Fig. 3).

Stable isotope analysis. Oxygen isotopic composition (δ¹⁸O) of stalagmite samples was analysed at three laboratories—Universität Innsbruck, Austria (the top 12.4 cm of sample SB-14, ~2,400 subsamples), Nanjing Normal University, China (133, 585, 140 and 155 subsamples from SB-12, SB-14, SB-32 and SB-58, respectively), and Xi'an Jiaotong University (80 subsamples from D8). Results are reported in per mil (‰), relative to the Vienna Pee Dee Belemnite (VPDB) standard. We obtained a total of ~3,360 stable isotope data (Supplementary Table 1). The techniques used in Universität Innsbruck are described in ref. 53. Stable isotope samples were micromilled perpendicularly to the extension axes of the stalagmites at 0.05 to 0.1 mm increments and analysed using an on-line carbonate preparation system (Gasbench II) interfaced with an isotope ratio mass spectrometer (Delta^{plus}XL). The long-term reproducibility is ~0.08‰ (1σ). The stable isotope measurements in Nanjing Normal University were made on a Thermo-Finnigan MAT-253 mass spectrometer fitted with a Kiel Carbonate Device III. Stable isotope samples were milled using carbide dental burs ranging in size from 0.3 to 0.5 mm along the central growth axis of stalagmites. Samples were calibrated against the NBS-19 standard. Standard measurements have an analytical precision of typically 0.08‰ (1σ). The measurements in Xi'an Jiaotong University were made on a Thermo-Finnigan MAT-253 mass spectrometer fitted with a Kiel Carbonate Device IV. Stable isotope samples were micromilled perpendicularly to the extension axes of the stalagmite D8 at 0.05 to 0.1 mm increments. Duplicate measurements of NBS19 and TBT1 standards show a long-term reproducibility of ~0.1‰ or better (1σ).

Composite Chinese cave δ¹⁸O record over the last 640 kyr. The composite AM δ¹⁸O record over the last 640 kyr is based on previously published data (<384 kyr BP)^{2–4,22,54,55} and new data from four Sanbao stalagmite records (≥384 kyr BP) from this study (SB-12, SB-14, SB-32 and SB-58). The composite δ¹⁸O record is constructed based on two criteria to select temporally-overlapped records—the dating control and the temporal resolution (see details in Supplementary Table 1). The new δ¹⁸O record from Dongge Cave (D8) replaces the previously published records between 217.2 and 225.3 kyr BP, as the age control of stalagmite D8 is superior for this time period. The overall average temporal-resolution of the record is 85 years. The lowest resolution (~500 years) is in the interval of the previous record between ~305 and 260 kyr BP, and the rest of time periods have resolution better than 200 years (Supplementary Table 1).

Normalization and detrending. In order to remove the orbital insolation component from the composite AM δ¹⁸O data, we used two different methods, z-standard and principal components analysis (PCA). For the z-standard method, we first converted the composite δ¹⁸O data to z-standard by using the mean and standard deviation of each data set (that is, zero-mean normalization). An average data resolution of 200 years was used to allow point-to-point alignment. The detrended δ¹⁸O data (Δδ¹⁸O) was obtained by subtracting 21 July insolation at 65°N (refs 3, 6)

from the δ¹⁸O record after standardization and equal spacing to the 200-year interval. Insolation is normalized to be equal in magnitude to δ¹⁸O and assigned an opposite algebraic sign: $\Delta\delta^{18}\text{O} = \delta^{18}\text{O} - (-\text{NHSI}_{\text{norm}}) = \delta^{18}\text{O} + \text{NHSI}_{\text{norm}}$. For the PCA method, we also converted the composite AM δ¹⁸O data in 200-year time-steps to synthesize the insolation in the same time-steps. We then standardized the common principal components obtained from the insolation time series and the original AM records. The PCAs are computed by using SPSS 15.0. The difference between the AM record and the common principal components was then used to characterize the millennial-scale variability. In addition to 21 July insolation, we also calculated the detrended results by using different insolation curves (21 June, 6 July and 6 August insolation at 65°N), in order to test the sensitivity. We found that the results are similar (Extended Data Figs 5 and 6). The methods and results used here are essentially similar to those described in ref. 8. The Δδ¹⁸O record obtained by the z-standard method is used in the figures for comparison.

The method used for detrending of the Antarctic δD record from EDC ice cores (Extended Data Fig. 8) was modified from that described in ref. 35. We first defined the long-term trend from the δD record by binning the combined data using the mean over 0.1-kyr intervals on 6-kyr windows on each data point. The window length of 6 kyr was chosen to effectively remove the glacial–interglacial trend³⁵. Sensitivity tests through varying the length of the 6-kyr window by a factor of two (that is, between 3 and 12 kyr) show no substantial impact on the spectral results. Second, we obtained the detrended record (ΔδD) by subtracting the long-term trend from the δD record. Our results are similar to those described in ref. 35.

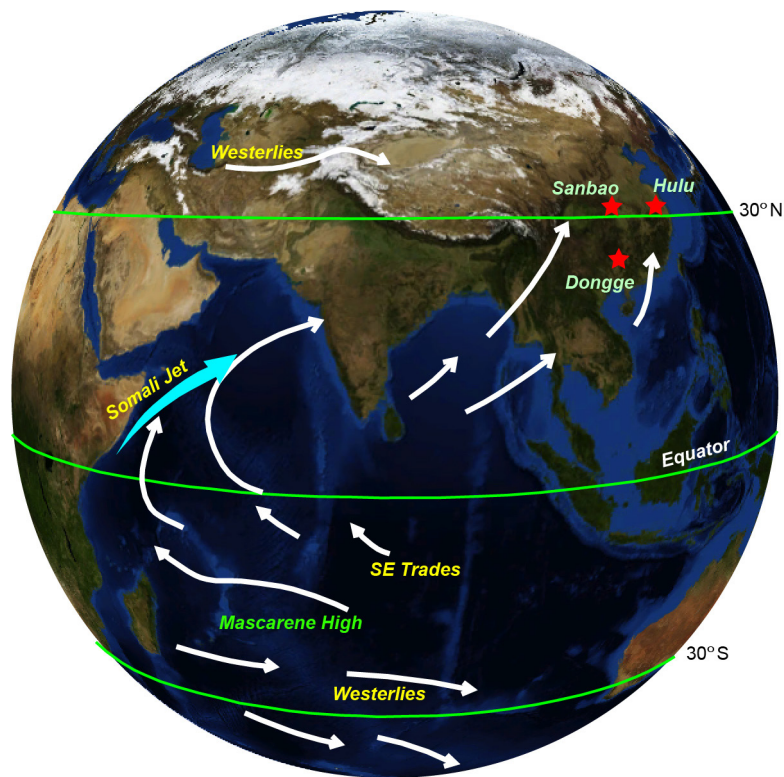
Spectrum analysis. In order to identify periodic components in the spectrum of our AM Δδ¹⁸O record, we applied the spectral analysis following the Blackman–Tukey method⁵⁶ using the ARAND software package. The following parameters were used to optimize bias/variance properties of spectrum estimates: number of lags = 1,200 (~1/3 length of record) and samples per analysis = 2,000. The results show that the Δδ¹⁸O records have significant power at ~23 kyr, weaker power at ~41 kyr and insignificant power at ~100 kyr (Fig. 4, Extended Data Figs 5 and 6).

Cross-spectrum analysis results are obtained between AM δ¹⁸O, AM Δδ¹⁸O, Antarctic ΔδD, and insolation records over the last 640 kyr using the ARAND software package. Detailed methods are described in ref. 56. The coherency spectra are compared with the 80% non-zero coherency level, resulting in a dominant cyclicity of ~23 kyr for all the analysis results (Fig. 4, Extended Data Figs 5, 6 and 8). However, the deviation of the phase spectrum from the zero-phase line suggests that the detrended AM records may have different phases relative to insolation at precession bands (~23 kyr), depending on the time of insolation used in the detrending process. If 21 July insolation is used for the detrending, as suggested by both empirical³ and theoretical⁶ studies, the millennial variability in the detrended AM record (Δδ¹⁸O record) is then nearly anti-phase with 21 June insolation (~180° ± 10° or ±1 kyr). In addition, because absolute ²³⁰Th dating errors increase progressively with sample age, we also tested all the spectrum analysis results by using the AM record only for the last 480 kyr BP when the dating errors are smaller. In all cases, we essentially obtained the same analysis results.

Chronologies. We use original chronologies of other climate records in most cases, including the Greenland ice core record (GICC05 chronology)⁴⁹, Antarctic ice core CO₂, CH₄ and δ¹⁸O_{atm}, dust and temperature records (EDC3 chronology)⁴⁶, the composite CO₂ record (AICC2012 chronology)²⁰, and the stacked benthic δ¹⁸O or the composite sea level record (LR04 chronology)^{17,57}. However, in order to compare our AM record with marine IRD and benthic δ¹⁸O records of ODP 980 and U1314 cores from the North Atlantic in Fig. 2, we have tuned their chronologies to our composite AM record through the correlation strategy as described in the main text: that is, by simple shifting of original chronologies to align the IRD events with WMLs as depicted by the grey bars in Fig. 2. In addition, we also tuned EDC ice core chronologies around the MIS 4/3 and 5.2/5.1 transitions in Fig. 2 by synchronizing the abrupt AM change with the CH₄ jump as described in the text. Amounts of age shifts from the original chronologies of the marine and ice core records are described in the figure legend.

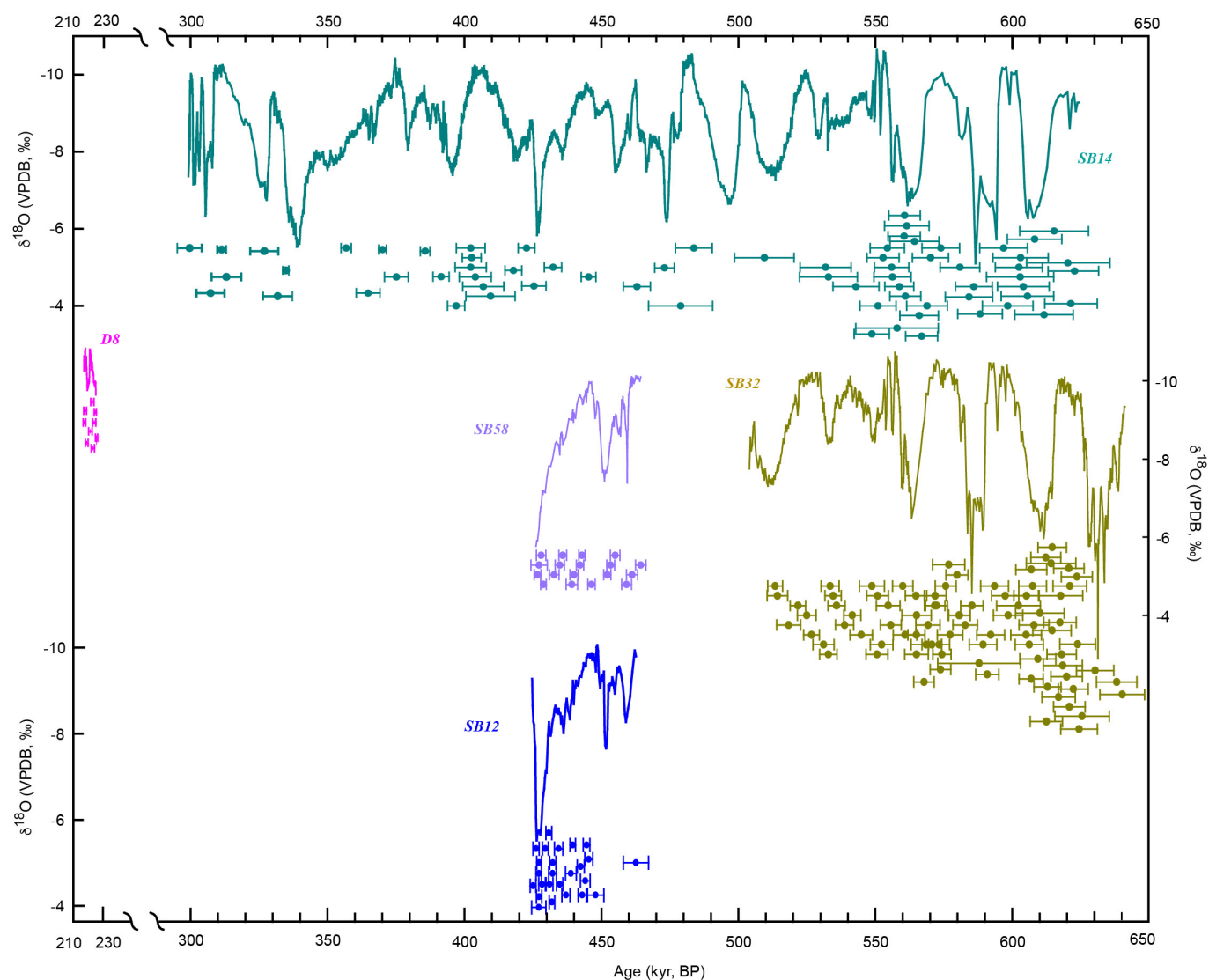
- Edwards, R. L., Chen, J. H. & Wasserburg, G. J. ²³⁸U–²³⁴U–²³⁰Th–²³²Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1987).
- Cheng, H. et al. The half-lives of U-234 and Th-230. *Chem. Geol.* **169**, 17–33 (2000).
- Spötl, C. & Vennemann, T. W. Continuous-flow isotope ratio mass spectrometric analysis of carbonate minerals. *Rapid Commun. Mass Spectrom.* **17**, 1004–1006 (2003).
- Dykoski, C. A. et al. A high resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China. *Earth Planet. Sci. Lett.* **233**, 71–86 (2005).

55. Kelly, M. J. *et al.* High resolution characterization of the Asian Monsoon between 146,000 and 99,000 years B.P. from Dongge Cave, China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **236**, 20–38 (2006).
56. Howell, P., Pisias, N., Ballance, J., Baughman, J. & Ochs, L. ARAND time-series analysis software (Brown Univ., 1997); available at <http://www.ncdc.noaa.gov/paleo/softlib/arand/arand.html>.
57. Lisiecki, L. E. & Raymo, M. E. A. Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**, PA1003 (2005).
58. Hendy, C. H. The isotope geochemistry of speleothems: I. The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as paleoclimate indicators. *Geochim. Cosmochim. Acta* **35**, 801–824 (1971).
59. Lambert, F. *et al.* Dust–climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**, 616–619 (2008).
60. Fleitmann, D. *et al.* Holocene forcing of the Indian monsoon recorded in a stalagmite from Southern Oman. *Science* **300**, 1737–1739 (2003).
61. Weldeab, S., Lea, D. W., Schneider, R. R. & Andersen, N. 155,000 years of West African monsoon and ocean thermal evolution. *Science* **316**, 1303–1307 (2007).
62. Holmgren, K. *et al.* Persistent millennial-scale climatic variability over the past 25,000 years in Southern Africa. *Quat. Sci. Rev.* **22**, 2311–2326 (2003).



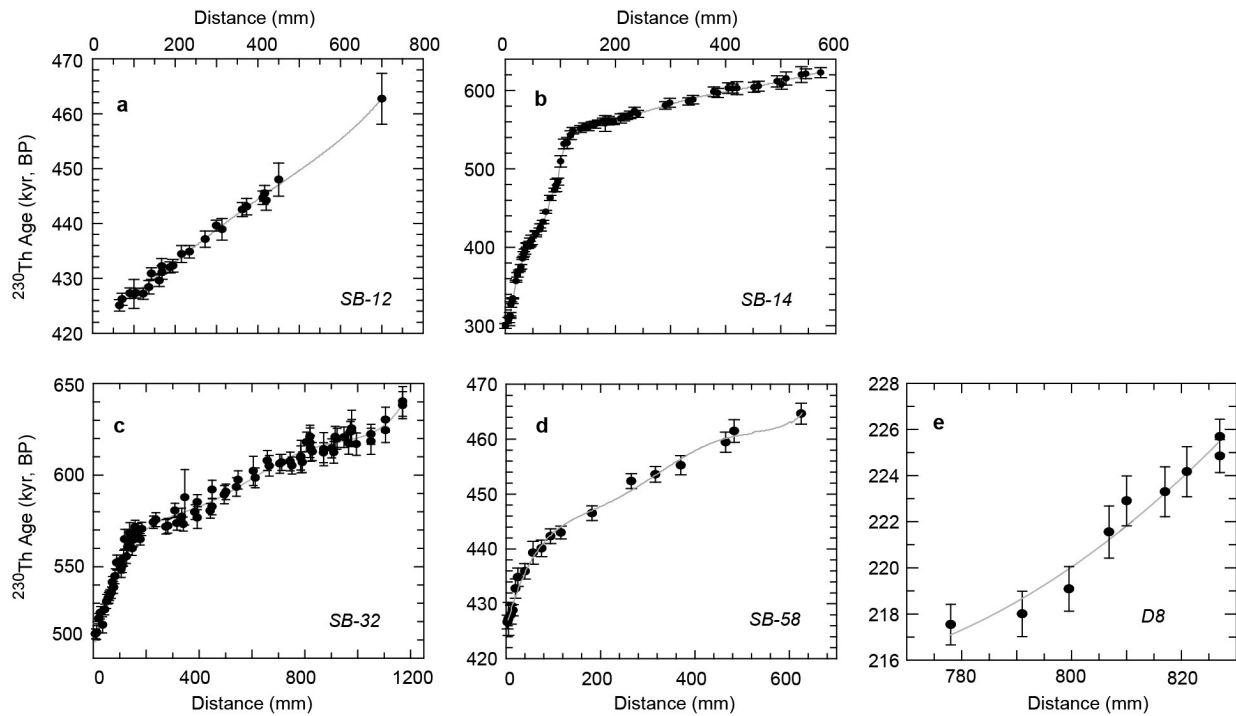
Extended Data Figure 1 | Schematic map of the vast Asian summer monsoon system. Arrows depict wind directions, yellow labels show wind names. The Mascarene High is a high pressure system near the Mascarene Islands. Stars indicate Sanbao (31° 40' N, 110° 26' E), Hulu (32° 30' N,

119° 10' E) and Dongge (25° 17' N, 108° 5' E) caves. The AM composite record is constructed from speleothem $\delta^{18}\text{O}$ records from these three caves. The map was constructed using NASA's World Wind program (<http://worldwind.arc.nasa.gov/java/>).



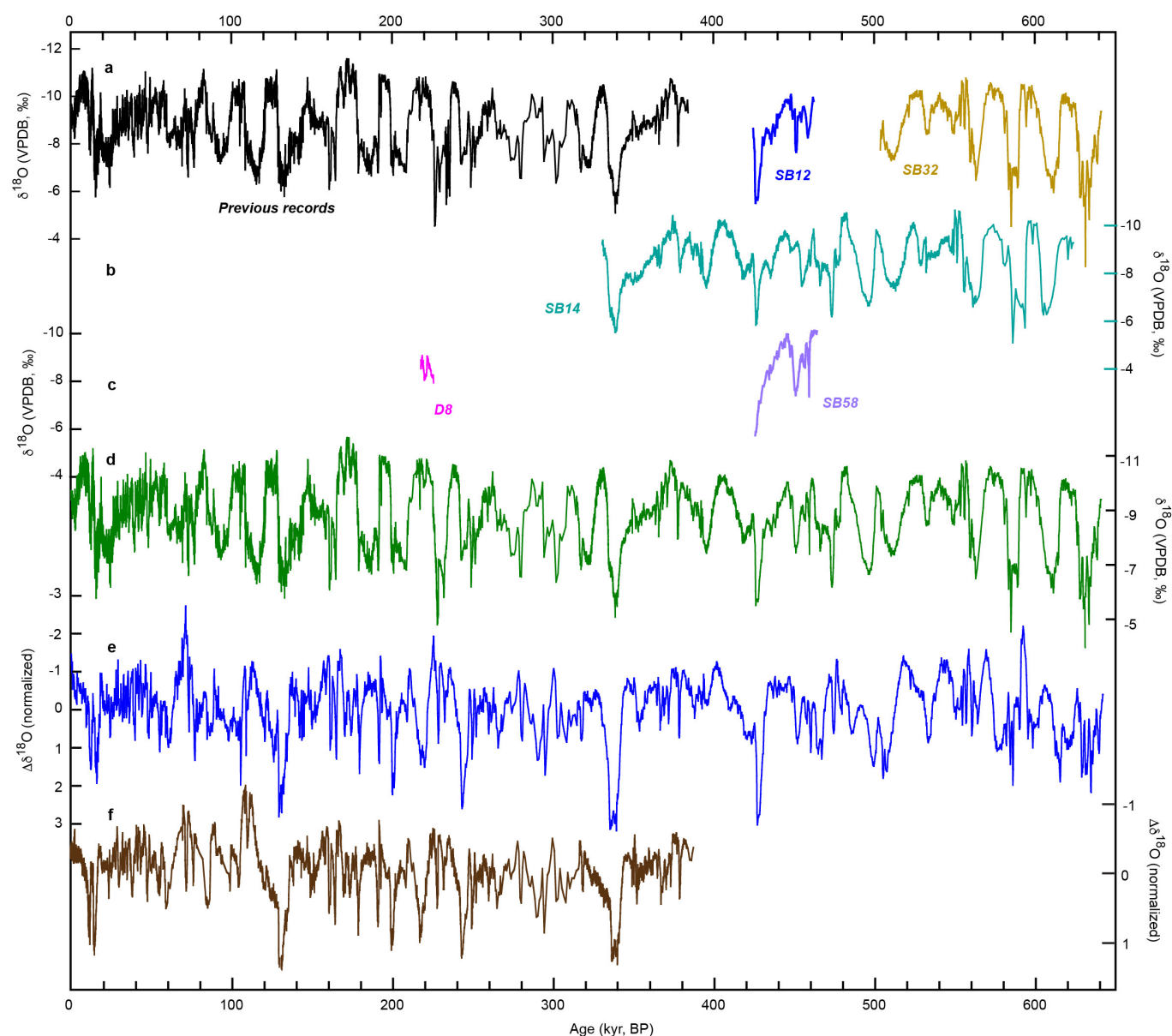
Extended Data Figure 2 | New speleothem records from China. Four new stalagmite $\delta^{18}\text{O}$ records used in this study are from Sanbao Cave, Hubei, China (labelled by sample numbers SB-12, SB-14, SB-32 and SB-58), and one is from Dongge Cave, Guizhou, China (labelled by sample number D8) (Fig. 1). Error bars indicate ^{230}Th ages and errors

(2σ). The ^{230}Th dating method is described in ref. 5 and the dating results are listed in Supplementary Table 1. A high degree of similarity among the coeval portions of different $\delta^{18}\text{O}$ records (the replication test^{2,58}) demonstrates that kinetic factors and water/rock interactions do not have a substantial effect on the speleothem $\delta^{18}\text{O}$ values.



Extended Data Figure 3 | Stalagmite age models. a–e, Age models are shown for five stalagmites: SB-12 (a), SB-14 (b), SB-32 (c), SB-58 (d) from Sanbao Cave, and D8 (e) from Dongge Cave. The chronology of the upper part of SB-14 is established by linear interpolation between

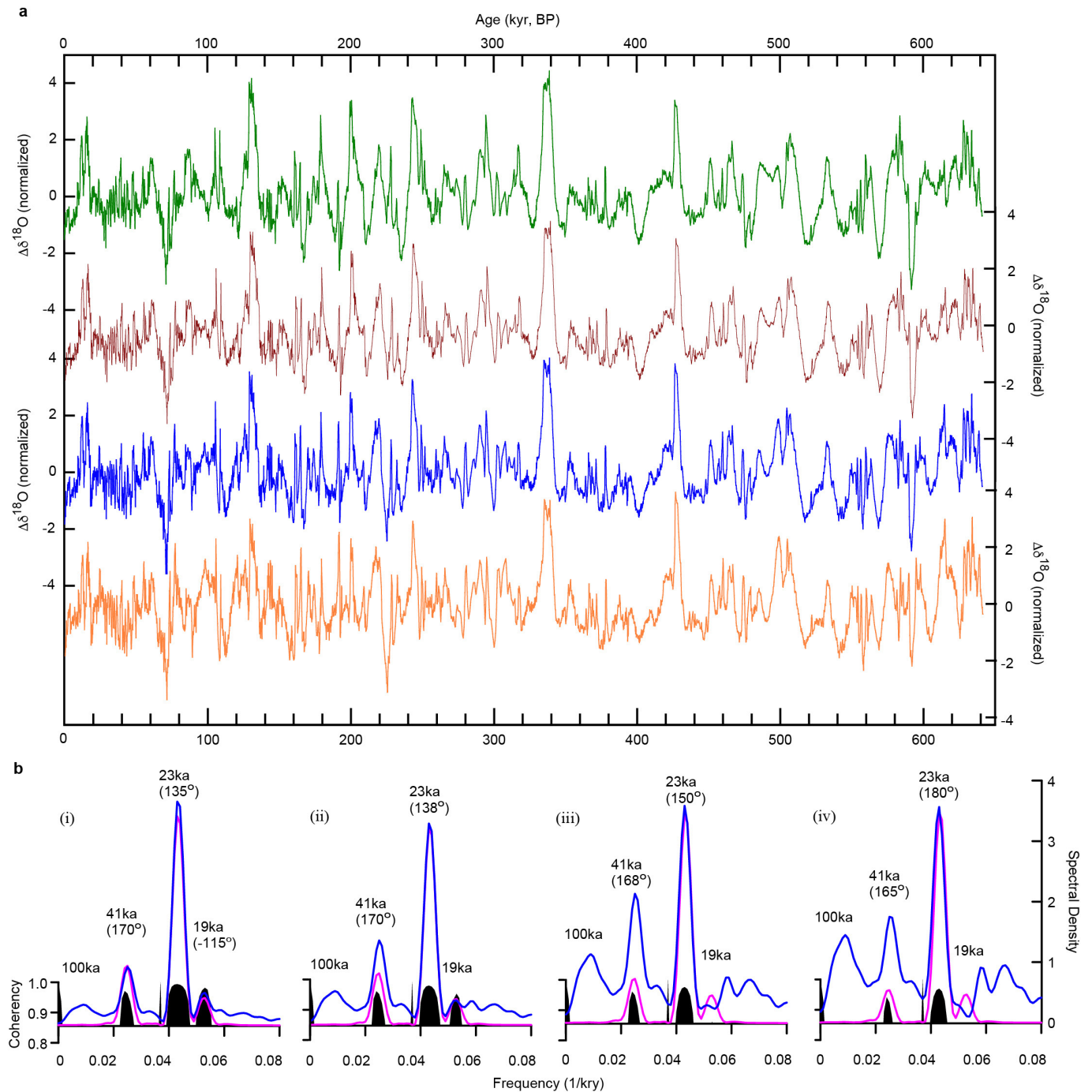
successive ^{230}Th dates, whereas chronologies for other samples are based on polynomial fitting of ^{230}Th dates. The vertical error bars depict errors (2σ) of ^{230}Th dates.



Extended Data Figure 4 | AM records over the past 640 kyr BP.

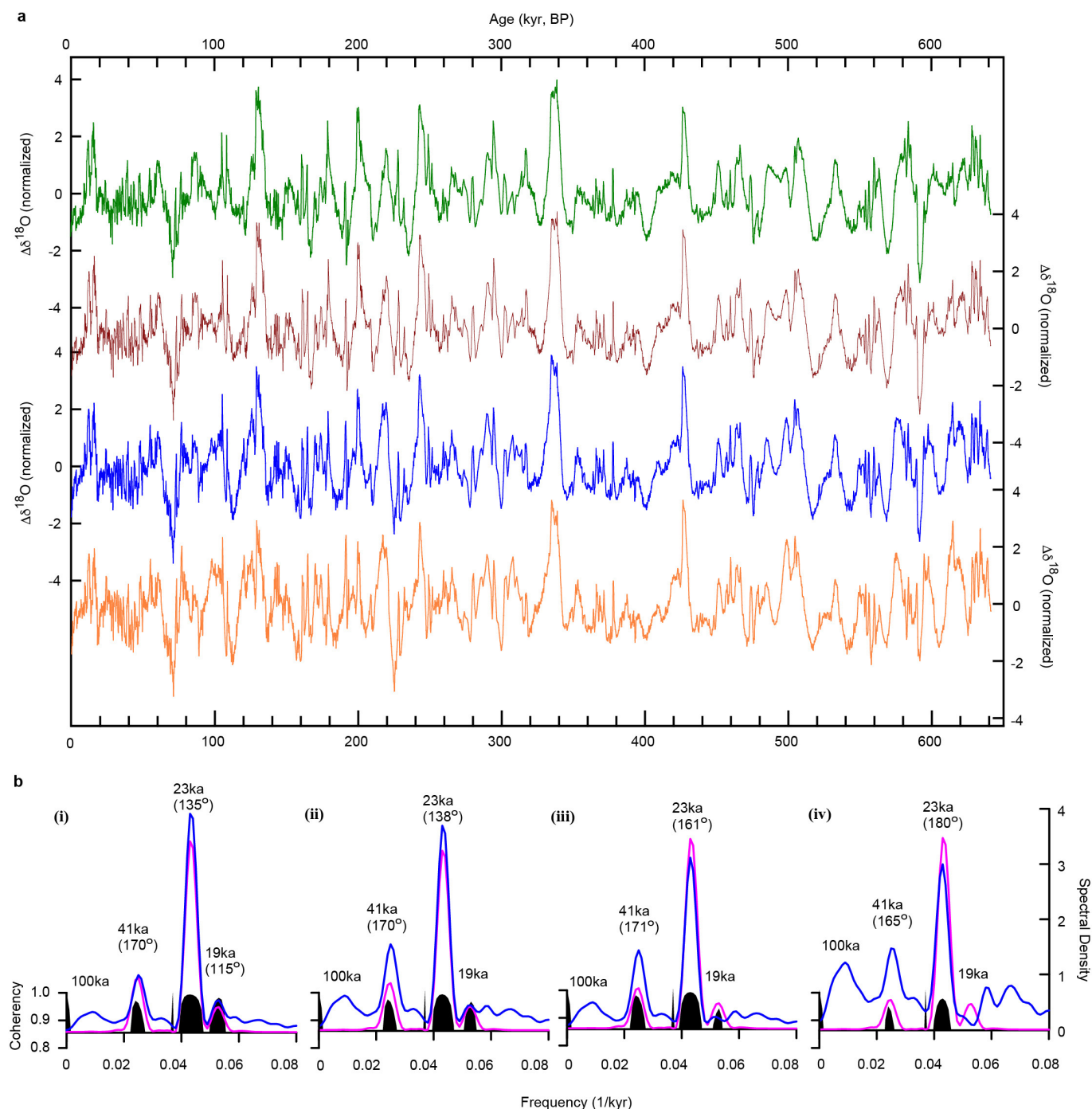
a, Previously published $\delta^{18}\text{O}$ records over the past 384 kyr BP (black) from Hulu, Dongge and Sanbao caves^{2-4,22,54,55}, and new SB-12 (blue) and SB-32 (olive) $\delta^{18}\text{O}$ records. **b**, New SB-14 $\delta^{18}\text{O}$ record. **c**, New SB-58 (purple) and D-8 (pink) $\delta^{18}\text{O}$ records. **d**, Composite AM $\delta^{18}\text{O}$ record over the past 640 kyr BP. The record is constructed from previous data (<384 kyr BP) and new data from four stalagmites from Sanbao Cave (≥ 384 kyr BP)

(see details in Methods and Supplementary Table 1). **e**, Detrended AM record ($\Delta\delta^{18}\text{O}$) is obtained by using the z-standard method to remove the insolation component (21 July insolation at 65°N) (see Methods). **f**, Detrended AM result from ref. 8, which is essentially identical to our results. Minor differences exist because the AM $\delta^{18}\text{O}$ records used in ref. 8 are slightly different.

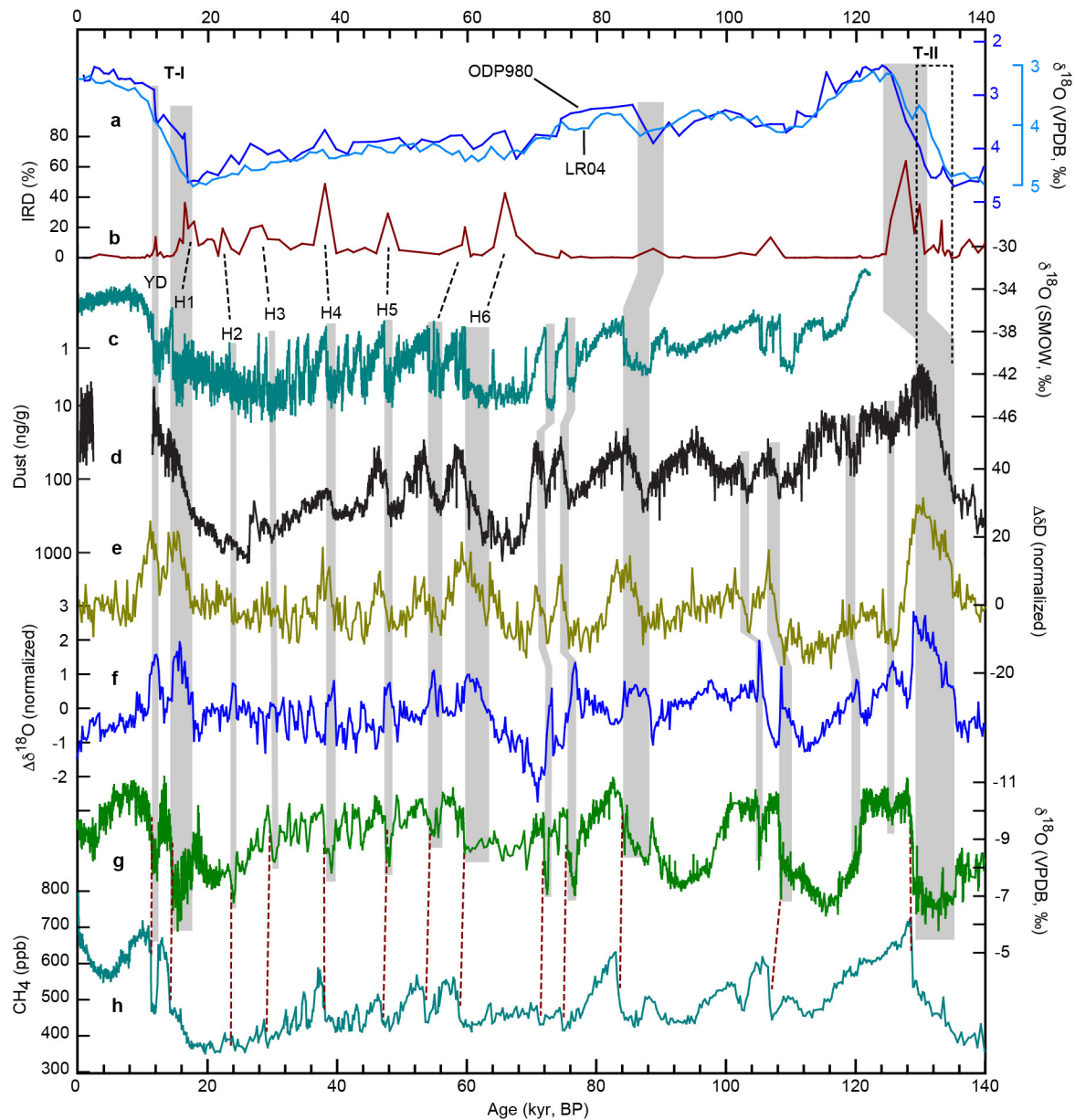


Extended Data Figure 5 | The AM $\Delta\delta^{18}\text{O}$ record over the past 640 kyr BP obtained by the z-standard method and cross-spectral comparison with insolation. **a**, Comparison among different AM $\Delta\delta^{18}\text{O}$ records detrended by subtracting 21 June (green), 6 July (brown), 21 July (blue) and 6 August (orange) insolation respectively, using the z-standard method (see Methods). **b**, Cross-spectral comparisons between the insolation for detrending (pink) and $\Delta\delta^{18}\text{O}$ records (blue) detrended by

21 June (i), 6 July (ii), 21 July (iii) and 6 August (iv) insolation respectively, using the z-standard method. Numbers in parentheses show the phase differences in degrees between insolation and $\Delta\delta^{18}\text{O}$ record. In all cases, the most significant power in the $\Delta\delta^{18}\text{O}$ record is in the precession band (~ 23 kyr). The ~ 41 -kyr power is also present, but is relatively weak. The ~ 100 -kyr power is insignificant.

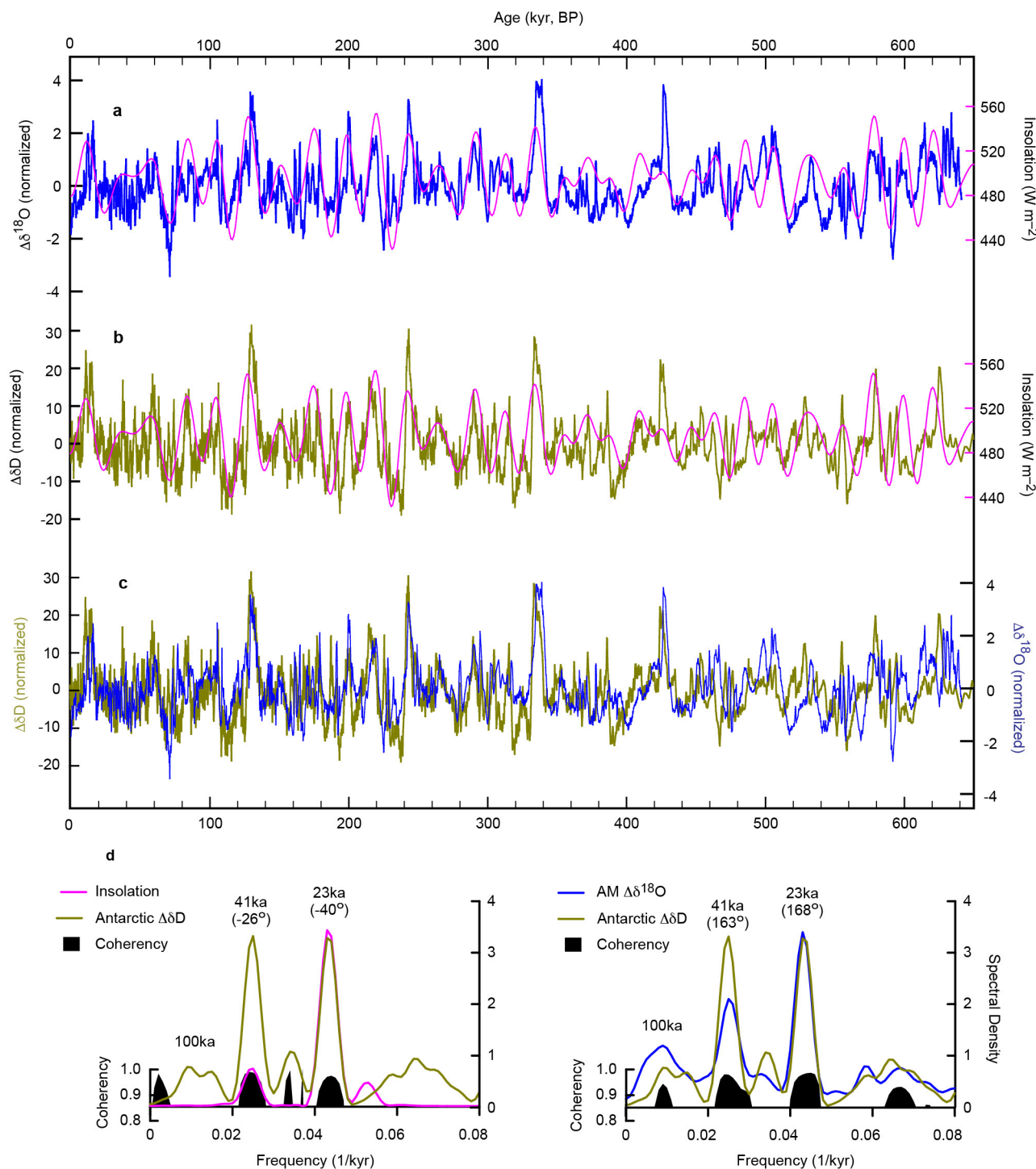


Extended Data Figure 6 | The AM $\Delta\delta^{18}\text{O}$ record over the past 640 kyr BP obtained by the principal component analysis method and cross-spectral analysis with insolation. a, b, As Extended Data Fig. 5, except the principal component analysis method was used, instead of the z-standard method.



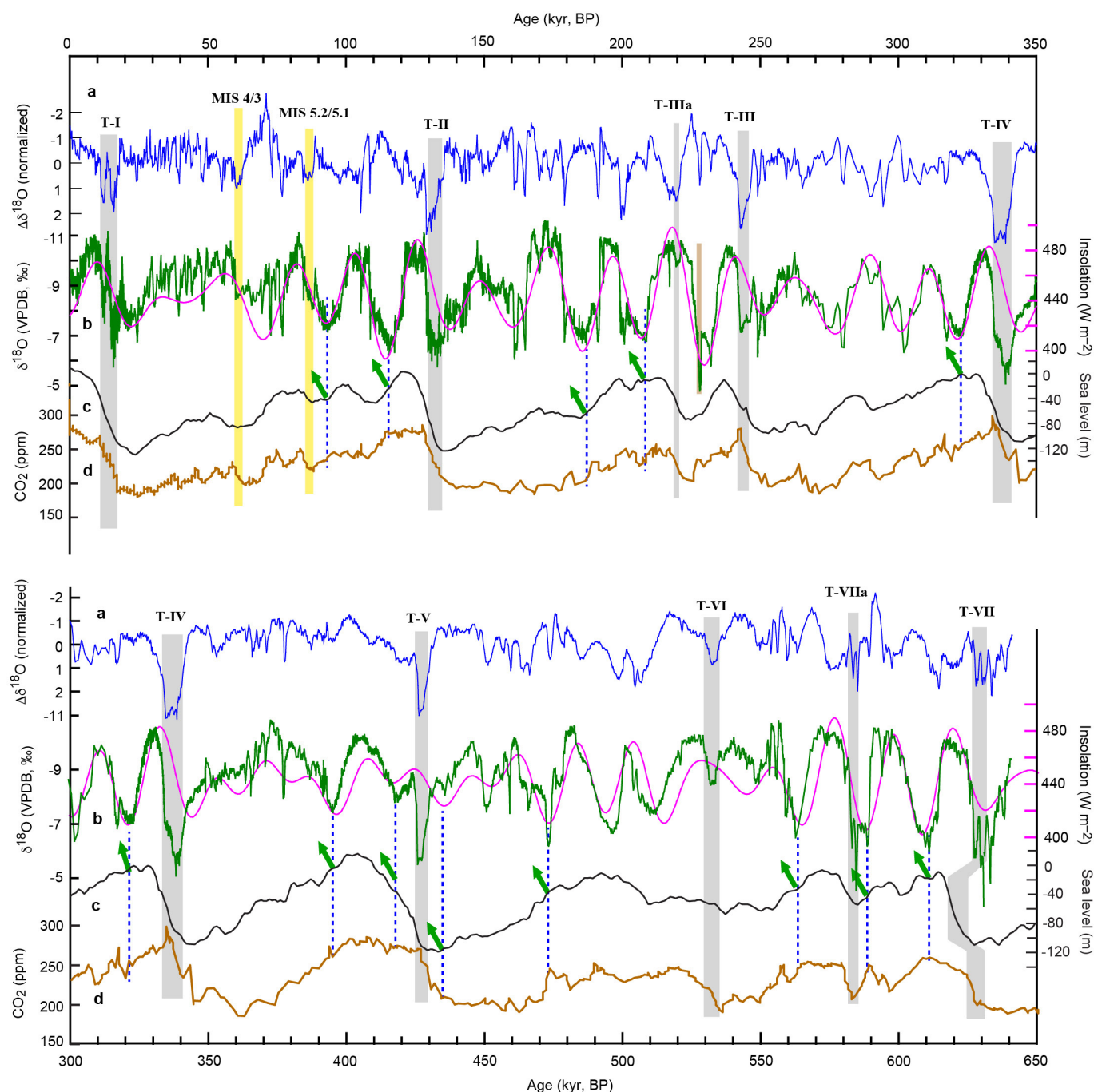
Extended Data Figure 7 | Comparison of millennial-scale climate events over the past 140 kyr. **a**, North Atlantic ODP980 (dark blue)⁷ and LR04 (light blue)⁵⁷ benthic $\delta^{18}\text{O}$ records. **b**, ODP980 IRD records⁷. We correlate the IRD event around T-II to the WMI with similar duration (depicted by grey bar), which is consistent with the obvious offset of its $\delta^{18}\text{O}$ shift related to that in LR04 record (dashed bar). YD and H1 to H6 indicate the Younger Dryas event, and Heinrich Stadial events 1 to 6, respectively. **c**, Greenland ice core (NGRIP) $\delta^{18}\text{O}$ record⁴⁹. **d**, Antarctic ice core (EDC) dust record⁵⁹. **e**, Detrended Antarctic δD record¹⁸ ($\Delta\delta\text{D}$), using a method modified from ref. 35. **f**, AM millennial variability ($\Delta\delta^{18}\text{O}$, detrended

from the composite AM $\delta^{18}\text{O}$ record by subtracting 21 July insolation at 65°N). **g**, Composite AM $\delta^{18}\text{O}$ record. **h**, EDC CH_4 record¹⁹. Vertical grey bars indicate major weak AM intervals (WMIs) and corresponding events (increase of temperature and decrease of dust flux in Antarctica, cold events in Greenland and IRD events in the North Atlantic Ocean). Dashed lines depict correlations between abrupt AM intensification and CH_4 jump (brown) and between weak monsoon and IRD events (black). All ice core records are on their EDC3 chronology⁴⁶. Notably, the AM $\Delta\delta^{18}\text{O}$ and Antarctic $\Delta\delta\text{D}$ records show striking similarity, demonstrating a common millennial-scale variability (Fig. 3).



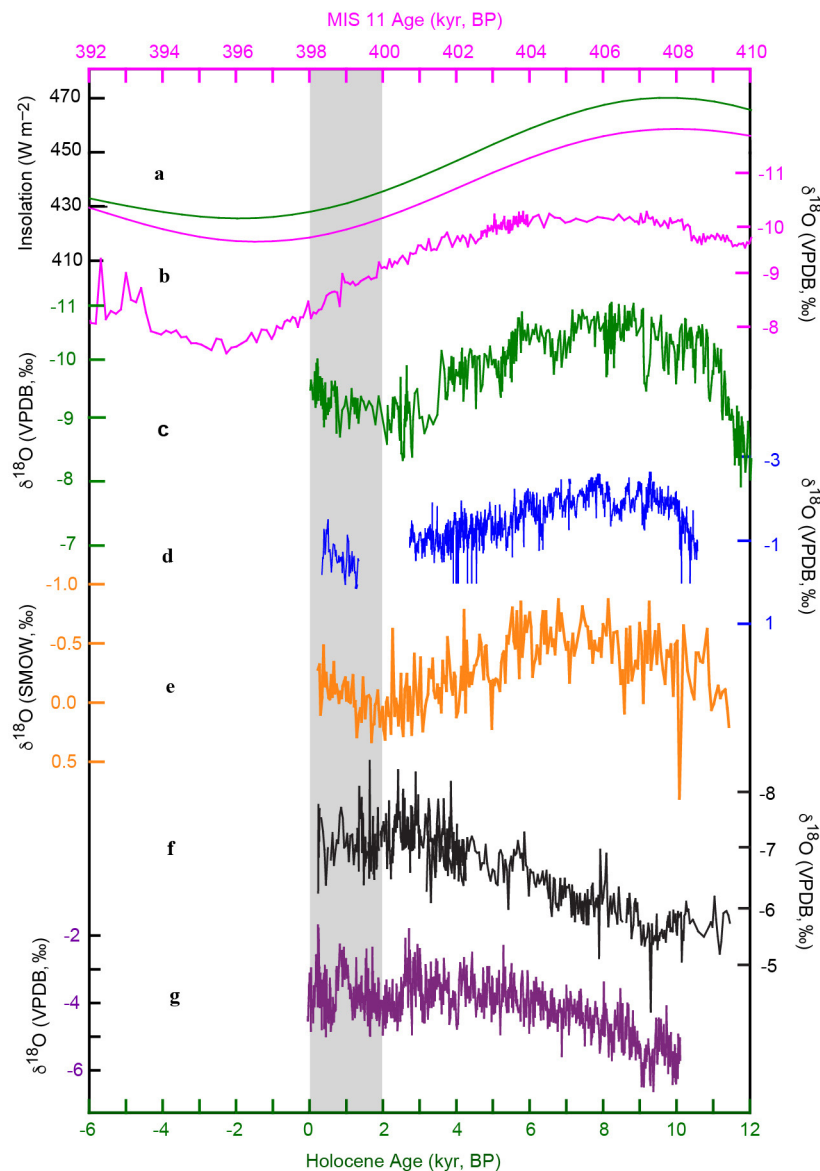
Extended Data Figure 8 | Comparison and cross-spectral analyses between insolation, AM $\Delta\delta^{18}\text{O}$ and Antarctic $\Delta\delta\text{D}$ records. **a**, AM $\Delta\delta^{18}\text{O}$ record (blue). **b**, Detrended Antarctic δD record¹⁸ ($\Delta\delta\text{D}$, olive), using a method modified from ref. 35. 21 June insolation at 65° N (ref. 45; pink) is plotted for comparison in **a** and **b**. **c**, Comparison between Antarctic $\Delta\delta\text{D}$ (olive) and AM $\Delta\delta^{18}\text{O}$ (blue) records. **d**, Cross-spectral analyses of the Antarctic $\Delta\delta\text{D}$ record (olive) with 21 July insolation at 65° N (pink)

(left), and with the AM $\Delta\delta^{18}\text{O}$ record (blue) (right), respectively. Numbers in parentheses show the phase differences in degrees. In all records, the precession cycle of ~23 kyr is significant. The phase of the Antarctic $\Delta\delta\text{D}$ record at precession band is close to 21 June insolation, and nearly anti-phased with the AM $\Delta\delta^{18}\text{O}$ record. The remarkable correlation between weak AM (positive $\Delta\delta^{18}\text{O}$ anomaly) and warm Antarctica (positive $\Delta\delta\text{D}$ anomaly) is evident (**c**).



Extended Data Figure 9 | Comparison of the AM variability with sea level (global ice volume) and atmospheric CO₂ changes. Upper panel, interval from 350 to 0 kyr BP; lower panel, interval from 650 to 300 kyr BP. In both panels: **a**, AM $\Delta\delta^{18}\text{O}$ record. **b**, AM $\delta^{18}\text{O}$ record (green) and 21 July insolation at 65° N (pink)⁴⁵. **c**, Composite sea level record¹⁷. **d**, Composite atmospheric CO₂ record²⁰. Grey bars show the timing of WMIs and associated terminations. Two yellow bars indicate the two millennial-scale positive anomalies (or WMIs), marking the ‘unfinished terminations’³⁰—the MIS 4/3 and 5.2/5.1 transitions (Fig. 2). For the

T-IIIa WMI we also indicate the correlation previously made by Cheng *et al.*⁴ with a beige bar. Although we consider this as a plausible alternative correlation, we prefer the new correlation presented in Figs 1, 3 and 5, and in this figure. The new correlation fits much better with the original chronologies of the ice core and marine records. In addition, the match of the adjacent high $\delta^{18}\text{O}$ AM anomalies and the ice rafted debris record⁷ is better with the new correlation. Some examples of initial AM rises around NHSI minima are depicted by green arrows and dashed lines, which do not appear to link directly to either global ice volume or CO₂ changes.



Extended Data Figure 10 | Comparison between the Holocene and MIS 11 on the basis of the insolation alignment. **a**, 21 July insolation at 65° N for the Holocene (green) and MIS 11 (pink)⁴⁵. **b**, **c**, Composite AM $\delta^{18}\text{O}$ records during MIS 11 and the Holocene, respectively. **d**, AM cave $\delta^{18}\text{O}$ record from the Indian monsoon domain⁶⁰. **e**, North African monsoon record (seawater $\delta^{18}\text{O}$ record from the marine sediment core, MD03-270, from the Gulf of Guinea)⁶¹. **f**, South American monsoon

record from Cueva del Tigre Perdido, northern Peru⁴⁰. **g**, South African monsoon record from Cold Air Cave, Makapansgat Valley, South Africa⁶². Vertical bar depicts the '2-kyr shift'. These records show '2-kyr shift' trends that are different from their trends exhibited in the middle to late Holocene interval. The monsoon '2-kyr shift' also appears to show an opposite inter-hemispheric pattern.

A combinatorial strategy for treating KRAS-mutant lung cancer

Eusebio Manchado¹, Susann Weissmueller^{1,2}, John P. Morris IV¹, Chi-Chao Chen^{1,3}, Ramona Wullenkord¹, Amaia Lujambio^{1,4}, Elisa de Stanchina⁵, John T. Poirier^{5,6}, Justin F. Gainor⁷, Ryan B. Corcoran⁷, Jeffrey A. Engelman⁷, Charles M. Rudin^{5,6}, Neal Rosen^{5,6} & Scott W. Lowe^{1,8}

Therapeutic targeting of KRAS-mutant lung adenocarcinoma represents a major goal of clinical oncology. KRAS itself has proved difficult to inhibit, and the effectiveness of agents that target key KRAS effectors has been thwarted by activation of compensatory or parallel pathways that limit their efficacy as single agents. Here we take a systematic approach towards identifying combination targets for trametinib, a MEK inhibitor approved by the US Food and Drug Administration, which acts downstream of KRAS to suppress signalling through the mitogen-activated protein kinase (MAPK) cascade. Informed by a short-hairpin RNA screen, we show that trametinib provokes a compensatory response involving the fibroblast growth factor receptor 1 (FGFR1) that leads to signalling rebound and adaptive drug resistance. As a consequence, genetic or pharmacological inhibition of FGFR1 in combination with trametinib enhances tumour cell death *in vitro* and *in vivo*. This compensatory response shows distinct specificities: it is dominated by FGFR1 in KRAS-mutant lung and pancreatic cancer cells, but is not activated or involves other mechanisms in KRAS wild-type lung and KRAS-mutant colon cancer cells. Importantly, KRAS-mutant lung cancer cells and patients' tumours treated with trametinib show an increase in FRS2 phosphorylation, a biomarker of FGFR activation; this increase is abolished by FGFR1 inhibition and correlates with sensitivity to trametinib and FGFR inhibitor combinations. These results demonstrate that FGFR1 can mediate adaptive resistance to trametinib and validate a combinatorial approach for treating KRAS-mutant lung cancer.

KRAS encodes a GTPase that couples growth factor signalling to the MAPK cascade and other effector pathways. Oncogenic KRAS mutations compromise its GTPase activity, leading to accumulation of KRAS in the active GTP-bound state and thereby to hyperactive signalling that initiates and maintains tumorigenesis¹. Owing to the high frequency of KRAS mutations in lung adenocarcinoma and other cancers, strategies to inhibit the KRAS protein or to exploit synthetic lethal interactions with a mutant KRAS gene have been widely pursued but have been fraught with technical challenges or produced inconsistent results^{2–7}. Conversely, strategies to target key RAS effectors including MAPK pathway components RAF, MEK, and ERK have been hindered by toxicities associated with their sustained inhibition and/or adaptive resistance mechanisms^{8–11}.

Screen to identify trametinib sensitizers

Hypothesizing that sustained MAPK inhibition is necessary, but not sufficient, for targeting KRAS-mutant cancers, we performed a pool-based short hairpin RNA (shRNA) screen to identify genes whose inhibition sensitizes KRAS-mutant lung cancer cells to the FDA-approved MEK inhibitor trametinib (Supplementary Table 1). A customized short-hairpin RNA (shRNA) library targeting the human kinome was introduced into the TRMPVIN vector that we previously optimized for negative selection screening^{12,13}. In this system, cassettes encoding a mir-30 shRNA linked to a dsRed fluorescent reporter are placed downstream of a tetracycline responsive promoter, enabling doxycycline dependent gene silencing and the facile tracking and/or sorting of shRNA expressing cells (Extended Data Fig. 1a)¹². This library was transduced into H23 KRAS^{G12C} mutant lung cancer cells expressing a

reverse-tet-transactivator (rtTA3). The transduced populations were then treated with doxycycline in the presence or absence of 25 nM trametinib, a dose that effectively inhibits ERK signalling without substantially affecting proliferation (Extended Data Fig. 1b–e). After ten population doublings, changes in shRNA representation were determined by sequencing of shRNAs amplified from dsRed-sorted cells (Extended Data Fig. 1b).

As expected, shRNAs targeting essential genes (*RPA1* and *CDK11A*) were strongly depleted in both vehicle and trametinib-treated cells, whereas the relative representation of neutral non-targeting control shRNAs (*Renilla* (*REN*)) remained unchanged (Fig. 1a and Extended Data Fig. 1f, g). Using selection criteria that required an average four-fold or greater depletion between conditions, we identified 64 shRNAs corresponding to 53 genes that were selectively depleted upon MEK inhibition in trametinib-treated compared with untreated cells (Fig. 1a and Extended Data Fig. 1h). Of these, shRNAs targeting the eight genes for which multiple shRNAs identified as hits were validated using cell competition assays in multiple KRAS-mutant lung lines. These studies identified *BRAF*, *CRAF*, *ERK2*, and *FGFR1* as the top candidates in our screen (Fig. 1b and Extended Data Fig. 2a).

Trametinib has superior pharmacological properties compared with other MEK inhibitors because it impairs feedback reactivation of ERK¹⁰. Still, the fact that MAPK components were identified as hits in our screen implied that pathway reactivation eventually occurs. Indeed, although trametinib stably inhibits ERK signalling at 48 h—a time where rebound occurs with other agents¹⁰—we observed an increase in phospho-ERK after 6–12 days of drug exposure (Fig. 1c). This rebound was reduced by subsequently increasing

¹Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ³Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, New York 10065, USA. ⁴Present address: Department of Oncological Sciences, Liver Cancer Program, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁵Department of Molecular Pharmacology and Chemistry, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁶Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁷Massachusetts General Hospital Cancer Center, Department of Medicine and Harvard Medical School, Boston, MA 02114, USA. ⁸Howard Hughes Medical Institute, New York, New York 10065, USA.

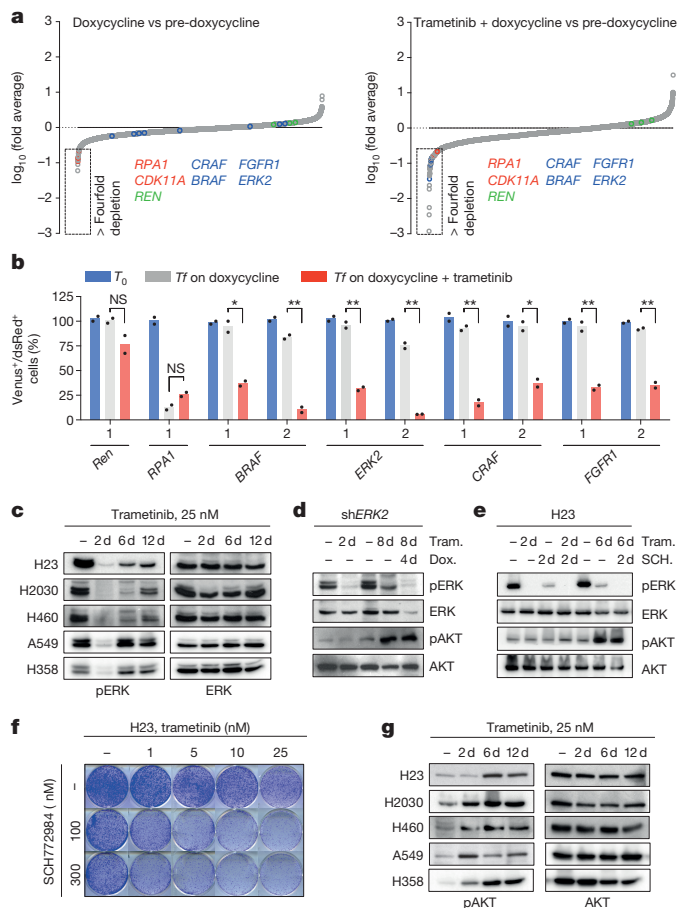


Figure 1 | Suppression of MAPK signalling effectors and FGFR1 sensitizes KRAS-mutant lung cells to trametinib. **a**, Relative abundance of each shRNA in the library in vehicle- or trametinib-treated H23 cells after ten population doublings on doxycycline. The mean of three (vehicle) and two (trametinib) replicates is plotted. Positive and negative controls included shRNAs targeting *RPA1* and *CDK11A* (red circles), and renilla (*REN*) luciferase (green circles). **b**, Quantification of fluorescent cells in competitive proliferation assays in H23 cells transduced with non-targeting control (*Ren*) or the indicated shRNAs. Data presented as mean ($n = 2$). * $P < 0.05$, ** $P < 0.01$ (unpaired two-tailed *t*-test). **c**, Immunoblot of KRAS-mutant lung cells treated with 25 nM trametinib for various times. **d**, Immunoblot of H23 cells transduced with a doxycycline-inducible shRNA targeting *ERK2* and treated with trametinib (Tram.; 25 nM) and doxycycline (Dox.) for the times shown. **e**, Immunoblot of H23 cells treated with trametinib (25 nM), SCH772984 (500 nM), or a combination for the times shown. **f**, Clonogenic assay of H23 cells treated with trametinib, ERK inhibitor SCH772984, or a combination as indicated ($n = 3$). **g**, Immunoblot of KRAS-mutant lung cancer cells treated with 25 nM trametinib for various times. For gel source data, see Supplementary Fig. 1.

the concentration of trametinib, indicating that it is MEK dependent (Extended Data Fig. 2b). Accordingly, inducible knockdown of *ERK2*, *CRAF*, and *BRAF* blocked ERK signalling rebound and reduced clonogenic growth after trametinib treatment (Fig. 1d and Extended Data Fig. 2c, d). Similar effects were observed in KRAS-mutant lung cancer cells treated with trametinib and the ERK inhibitor SCH772984 (Fig. 1e, f and Extended Data Fig. 3)¹⁴. These observations emphasize the marked dependency of KRAS-mutant tumours on the MAPK signalling pathway.

In agreement with other studies, KRAS-mutant cells treated with trametinib also displayed compensatory activation of the PI3K and JAK/STAT pathways as assessed by AKT and STAT3 phosphorylation, respectively (Fig. 1d, e, g and Extended Data Figs 2c, 3b and 4a)^{11,15}. Although the increase in STAT3 phosphorylation was

transient (Extended Data Fig. 4a), AKT phosphorylation was sustained (Fig. 1g). In contrast to their effects on ERK signalling rebound, genetic or pharmacological inhibitions of MAPK signalling had little effect on the trametinib-induced increase in pAKT (Fig. 1d, e and Extended Data Figs 2c and 3b). The activation of multiple signalling pathways after trametinib treatment probably reflects a relief in pleiotropic feedback mechanisms produced by hyperactive RAS signalling in KRAS-mutant cells^{8,9}.

FGFR1 mediates adaptive drug resistance

Several RTKs have been implicated in adaptive resistance to RAS pathway antagonist^{8,9,11,15–20}. The identification of *FGFR1* shRNAs as trametinib sensitizers raised the possibility that FGFR1 mediates MAPK and PI3K activation in trametinib-treated KRAS-mutant cells. In agreement, treatment of KRAS-mutant lung tumour cell lines with trametinib increased FGFR1 receptor and/or ligand expression together with FGFR pathway activation as assessed by an increase in phosphorylation of the FGFR adaptor protein FRS2 (Fig. 2a, b and Extended Data Figs 2b and 4b–e)²¹. In turn, FGFR1 activation correlated with an increase in the levels of RAS-GTP, phospho-AKT, and phospho-ERK (Fig. 2b and Extended Data Fig. 4e), which was prevented by FGFR1 knockdown (Fig. 2c). Accordingly, *FGFR1* shRNAs did not inhibit the proliferation of KRAS-mutant lung cancer, but displayed synergistic inhibitory effects when combined with trametinib (Fig. 2d, e and Extended Data Fig. 4f, g).

The combinatorial effects of FGFR1 inhibition and trametinib showed distinct specificities: for example, shRNAs targeting *FGFR1* or *FRS2*, but not those targeting *FGFR2* and *3*, sensitized KRAS-mutant lung cancer cells to trametinib (Fig. 2f and Extended Data Fig. 4h, i). By contrast, *FGFR1* knockdown had little impact on trametinib sensitivity in KRAS wild-type lung cancer cells (Fig. 2g). While *FGFR1* shRNAs synergized with trametinib in two KRAS-mutant pancreatic cancer cell lines, they showed little activity in trametinib-treated KRAS-mutant colorectal lines (Extended Data Fig. 5a). Importantly, this genotype and tissue specificity correlated with the ability of trametinib to trigger FRS2 phosphorylation when applied as a single agent (Extended Data Fig. 5b–d). Therefore, treatment of certain KRAS-mutant tumour types with trametinib induces a dependency on FGFR1 signalling that promotes adaptive drug resistance.

FGFR1 inhibition enhances trametinib effects

We next tested whether therapeutic strategies combining trametinib with an FGFR1 inhibitor could be effective in treating some KRAS-mutant lung cancers by combining trametinib with ponatinib, an FDA-approved multikinase inhibitor that inhibits FGFR1 and is being tested clinically for activity against FGFR1-amplified lung cancer (Extended Data Fig. 6a) (ref. 22 and <https://clinicaltrials.gov/show/NCT01935336>). Ponatinib had little effect on KRAS-mutant cells but countered the trametinib-induced increase in pFRS2, pERK, and pAKT, and synergized with trametinib in inhibiting cell proliferation (Fig. 3a–c, and Extended Data Fig. 6b–e). As observed in our genetic studies, this combination also showed combined activity in human KRAS-mutant pancreatic cancer cells and a *Kras*-mutant murine lung adenocarcinoma line (Extended Data Fig. 7a–c), but to a lesser extent in KRAS wild-type lung cancer cells or KRAS-mutant colon cancer cells (Fig. 3c). Although it remains possible that the synergistic effects of this combination involve the ability of ponatinib to target additional kinases, similar results were observed with two other chemically distinct FGFR inhibitors (Extended Data Figs 7d, e and 8a)^{23,24}. Importantly, sensitivity to the combination of trametinib and FGFR inhibition correlated with the degree of pFRS2 induction after trametinib treatment (Extended Data Fig. 8b, c).

While our genetic and pharmacological studies establish the importance of the MAPK pathway in adaptive resistance to trametinib,

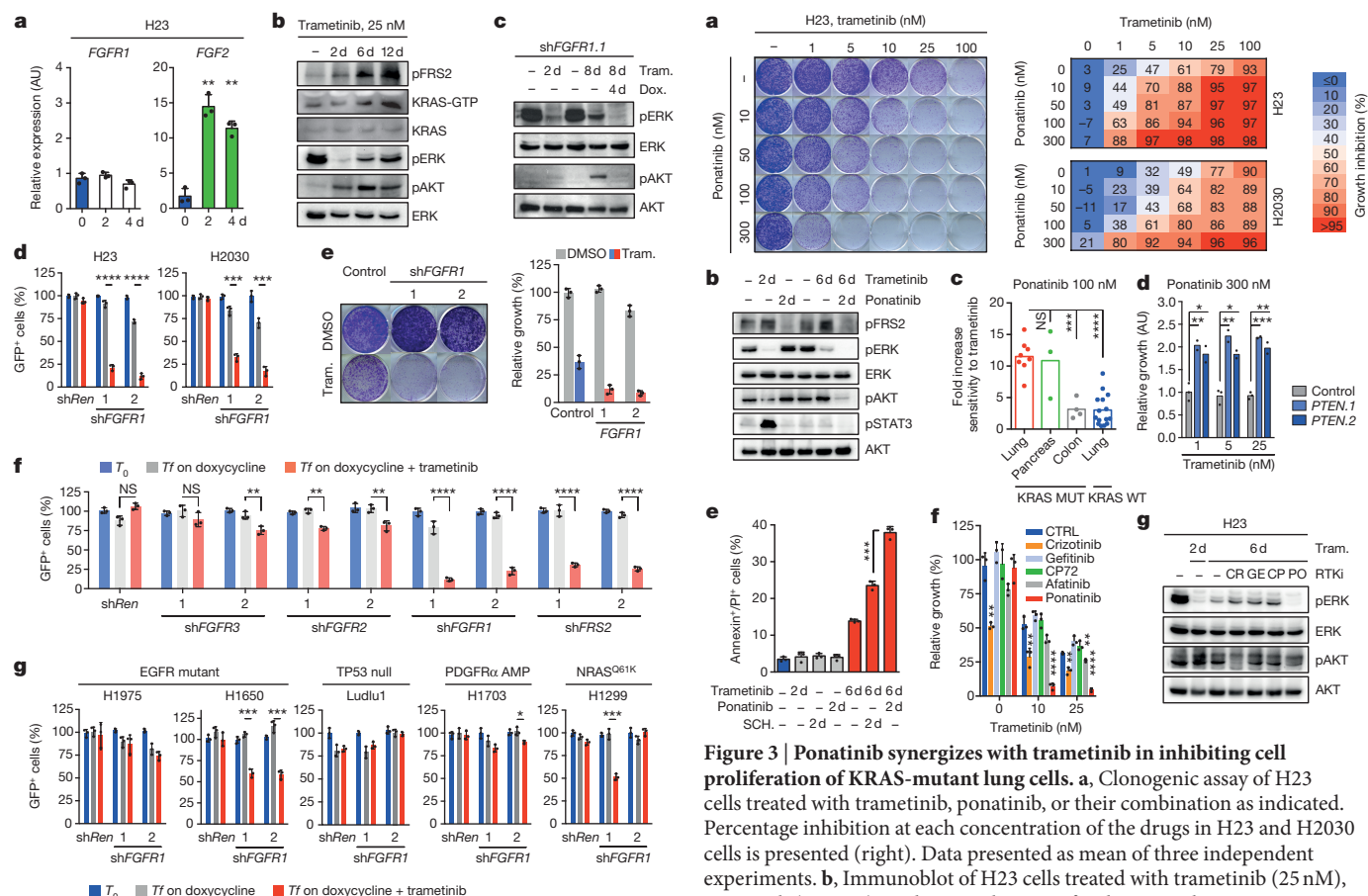


Figure 2 | Feedback activation of FGFR1 mediates adaptive resistance to trametinib in KRAS-mutant lung cells. **a**, Quantitative reverse transcription PCR (qRT-PCR) for *FGFR1* and *FGF2* in H23 cells treated with trametinib for the indicated times ($n = 3$). AU, arbitrary units. **b**, Immunoblot of H23 cells treated with 25 nM of trametinib for various times. **c**, Immunoblot of H23 cells transduced with a doxycycline-inducible shRNA targeting *FGFR1* and treated with trametinib (25 nM) and doxycycline for the times shown. **d**, Quantification of fluorescent cells in competitive proliferation assay in H23 and H2030 cells transduced with doxycycline-inducible non-targeting control (*Ren*) or *FGFR1* shRNAs ($n = 3$). **e**, Clonogenic assay of H23 cells transduced with *FGFR1* and non-targeting control shRNAs, and cultured with dimethylsulfoxide (DMSO) or trametinib (25 nM). Relative growth of DMSO- (grey bars) and trametinib-treated cells (blue and red bars) is shown (right) ($n = 3$). **f**, **g**, Quantification of fluorescent cells in competitive proliferation assays in H23 (**f**) and the indicated lung cancer cells (**g**) transduced with doxycycline-inducible non-targeting control (*Ren* (Renilla)) or the indicated shRNAs ($n = 3$). **a**, Paired two-tailed t -test. **d**, **f**, **g**, Unpaired two-tailed t -test. Data presented as mean \pm s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For gel source data, see Supplementary Fig. 1.

we reasoned that the compensatory increase in PI3K/AKT signalling also plays a role and that its inhibition by ponatinib contributes to the effects of this drug combination. Accordingly, PTEN knockdown, which can increase PI3K signalling independently of RTK activation, promoted partial resistance to the drug combination in KRAS-mutant H2030 cells. This effect was not observed in H460 cells, a KRAS-mutant line that also harbours an activating mutation in the p110 α catalytic subunit of PI3K (Fig. 3d and Extended Data Fig. 9a–c). Consistent with a role for PI3K signalling in promoting cell survival, co-treatment of H23 cells with trametinib and ponatinib triggered substantial apoptosis in a manner that was not observed after treatment with trametinib alone or in combination with an ERK inhibitor (Fig. 3e and Extended Data Fig. 9d). Thus, the combined ability of ponatinib to impact reactivation

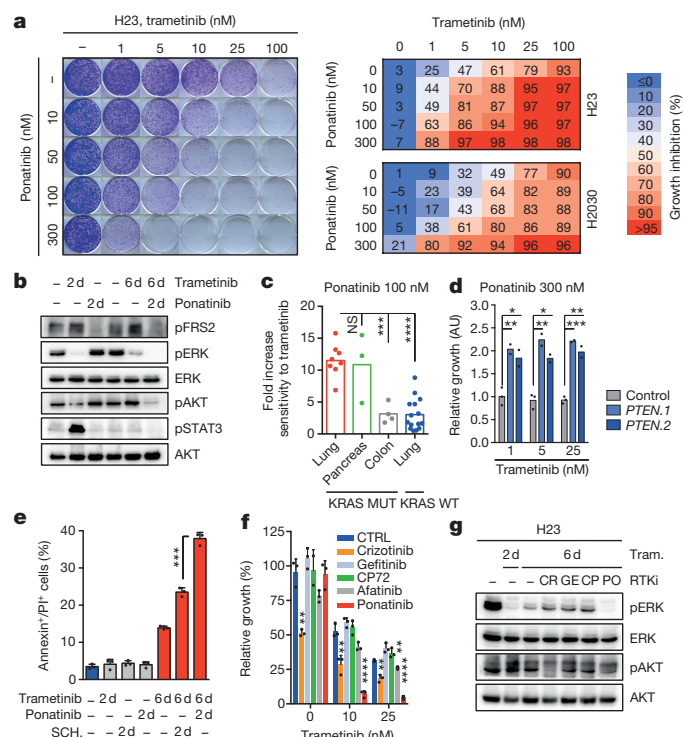


Figure 3 | Ponatinib synergizes with trametinib in inhibiting cell proliferation of KRAS-mutant lung cells. **a**, Clonogenic assay of H23 cells treated with trametinib, ponatinib, or their combination as indicated. Percentage inhibition at each concentration of the drugs in H23 and H2030 cells is presented (right). Data presented as mean of three independent experiments. **b**, Immunoblot of H23 cells treated with trametinib (25 nM), ponatinib (750 nM), or their combination for the times shown. **c**, Dot plot illustrating the sensitivity increase to trametinib after treatment with ponatinib (100 nM) in a panel of KRAS-mutant ($n = 15$) and KRAS wild-type ($n = 15$) cancer cell lines. Data presented as mean of two independent replicates. **d**, Quantification of the relative growth of H2030 cells transduced with *PTEN* and non-targeting control shRNAs, and treated with ponatinib (300 nM) in combination with trametinib (1, 5, and 25 nM). Data presented as mean of two independent replicates. **e**, Quantification of AnnexinV/PI double positive cells in H23 cells treated with trametinib (25 nM), ponatinib (300 nM), SCH7272984 (1 μ M), or their combination for the times shown ($n = 3$). **f**, Quantification of the relative growth of H23 cells treated with trametinib alone or in combination with 500 nM crizotinib, gefitinib, CP-724714, afatinib, or 300 nM ponatinib ($n = 3$). **g**, Immunoblot of H23 cells pre-treated with trametinib (25 nM) for 4 days, followed by treatment with trametinib (25 nM) alone or in combination with crizotinib (1 μ M), gefitinib (1 μ M), CP-724714 (1 μ M), and ponatinib (750 nM) for 2 days. **c**–**f**, Unpaired two-tailed t -test. Error bars, mean \pm s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For gel source data, see Supplementary Fig. 1. Source Data for Fig. 3 are available in the online version of the paper.

of the MAPK and PI3K pathways contributes to its combinatorial activity in KRAS-mutant lung cancer cells.

We also tested whether other RTKs known to be reactivated after MAPK inhibition contribute to adaptive resistance to trametinib in KRAS-mutant lung cancer cells^{8,9,11,16–20,25,26}. While trametinib treatment of H23 and H2030 cells increased MET and ERBB2 (but not EGFR) levels (Extended Data Fig. 9e–g), inhibitors targeting these kinases did not synergize with trametinib under the conditions tested (Fig. 3f and Extended Data Figs 9h and 10a, b). Consistent with previous reports¹¹, the dual EGFR/ERBB2 inhibitor afatinib also showed combinatorial activity with trametinib in some KRAS-mutant lung cancer lines, although in our hands less robustly than the trametinib and ponatinib combination (Fig. 3f and Extended Data Figs 9h and 10a, b). Accordingly, none of the agents tested prevented the rebound in ERK signalling after trametinib treatment (Fig. 3g and

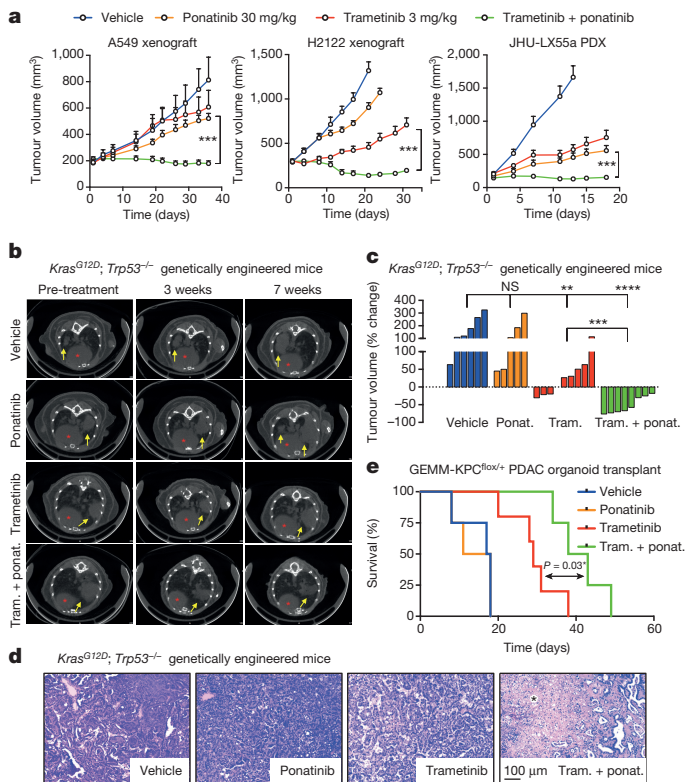


Figure 4 | Suppression of FGFR1 in combination with trametinib leads to regression of KRAS-mutant lung tumours. **a**, Tumour volumes of mice bearing A549 and H2122 xenografts, and JHU-LX55a patient-derived xenograft tumours and treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination for the indicated times. Error bars, mean \pm s.e.m. ($n \geq 6$ per treatment group). **b–d**, Representative micro-computed tomography images of the lungs of *Kras*^{G12D}; *Trp53*^{-/-} genetically engineered mice treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (ponat.) (30 mg/kg body weight), or both drugs in combination for 3 and 7 weeks. Lung tumours are indicated by yellow arrows; red asterisks mark the hearts (**b**). A waterfall representation of the response for each tumour after 3 weeks of treatment is shown ($n \geq 5$ per group) (**c**). Representative haematoxylin and eosin stains are shown. Black asterisk indicates necrosis (**d**). **e**, Kaplan–Meier survival analysis of mice bearing pancreatic tumours resulting from orthotopic transplantation of GEMM-KPC^{flac/+} PDAC organoids and treated as in **b** ($n \geq 4$ per group) (log-rank test). **a**, **c**, Unpaired two-tailed *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Source Data for Fig. 4 are available in the online version of the paper.

Extended Data Fig. 10c, d). Thus, reactivation of FGFR1 signalling is a prominent mechanism of adaptive resistance to trametinib in KRAS-mutant lung cancer cells.

In vivo effects of MEK/FGFR1 inhibition

We validated our *in vitro* results in KRAS-mutant lung cancer xenografts, a KRAS-mutant patient-derived xenograft, and a genetically engineered mouse model (GEMM) of *Kras*^{G12D}-induced lung adenocarcinoma that accurately resembles the human disease²⁷. A549 and H2122 xenografts harbouring tet-responsive *FGFR1*- or control-shRNAs were treated with doxycycline and a daily dose of 3 mg/kg body weight of trametinib when tumours reached ~ 150 mm³. While knockdown of FGFR1 or treatment with trametinib alone had only minor anti-tumour effects, the combination of FGFR1 knockdown with trametinib potently inhibited tumour growth and typically caused tumour regression (Extended Data Fig. 11a, b). Treatment of the xenografts, PDX and GEMM models with vehicle, trametinib, ponatinib, or the drug combination showed similar results, with only the combination producing marked tumour regressions despite no apparent toxicities (Fig. 4a–c

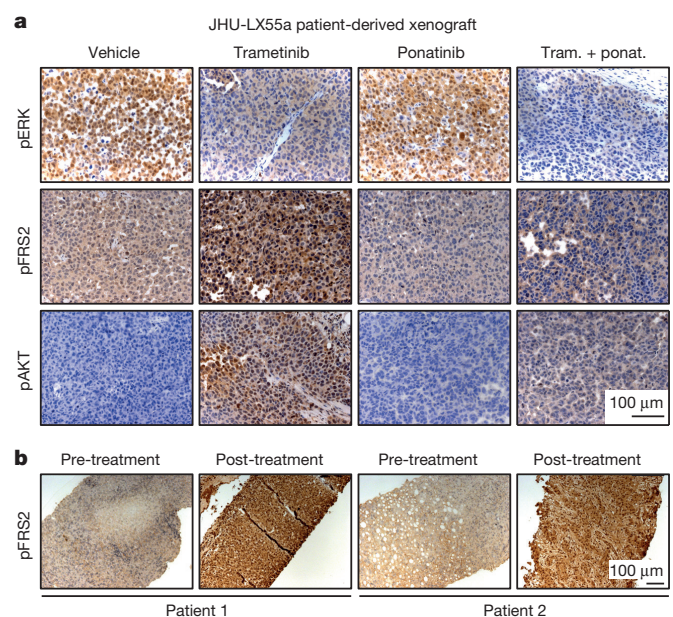


Figure 5 | Trametinib induces FGFR1 signalling in KRAS-mutant lung tumours. **a**, Tumour tissue from JHU-LX55a patient-derived xenografts treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination for 18 days was evaluated by IHC for phospho-FRS2, phospho-ERK, and phospho-AKT. **b**, Paired tumour biopsies from patients having KRAS-mutant lung adenocarcinomas (before and after treatment with the MEK inhibitor trametinib) were evaluated by IHC for phospho-FRS2.

and Extended Data Fig. 11c–e). Moreover, histological analysis of the residual tumour mass in GEMMs treated with the drug combination showed massive necrosis, an effect not seen with either agent alone (Fig. 4d). Similar results were observed in an organoid-based, transplantable model of *Kras*^{G12D}-driven pancreatic cancer, in which the drug combination produced marked cell death and significantly enhanced survival (Fig. 4e and Extended Data Fig. 11f).

We also examined the ability of trametinib to induce FGFR1 signalling in KRAS-mutant tumours. Consistent with *in vitro* results, a KRAS-mutant lung PDX model showed a concomitant increase in FRS2, ERK, and AKT phosphorylation after trametinib treatment—an effect that was cancelled by ponatinib (Fig. 5a and Extended Data Fig. 11g). Furthermore, FRS2 phosphorylation was dramatically increased after trametinib treatment in two patients with KRAS-mutant lung adenocarcinoma (Fig. 5b), indicating that the mechanism of adaptive resistance identified in our preclinical models is clinically relevant.

Discussion

In summary, by implementing a stringent approach for negative selection shRNA screening, we identified feedback activation of FGFR1 signalling as a prominent mechanism of adaptive resistance to the MEK inhibitor trametinib in KRAS-mutant lung cancer. The mechanism was specific: only shRNAs targeting *FGFR1*, but not other FGFR family members or other RTKs tested, conferred trametinib sensitivity, and only FGFR1 inhibition blocked compensatory reactivation of both ERK and AKT. In agreement, an unbiased ORF screen identified FGFR1, but not other RTKs, as sufficient to allow proliferation of KRAS-mutant colon cancer cells after KRAS suppression²⁸. In our hands, the synergistic effects of the trametinib/FGFR inhibitor combinations were largely restricted to KRAS-mutant lung and pancreatic cancer cells, but not KRAS wild-type lung or KRAS-mutant colon cancer cells. These results strongly associate sensitivity to the combination with the magnitude of FRS2 phosphorylation after trametinib treatment alone and provide a mechanistic foothold to predict and study cell line and tumour variability.

Our results provide strong mechanistic support for combining trametinib with FGFR1 inhibitors for treating KRAS-mutant lung cancer and pinpoint a biomarker that might eventually be used to identify other patients likely to benefit from this drug combination. Although careful attention to additive or synergistic toxicities will be required for the clinical implementation of these findings, it seems likely that targeting a specific RTK such as FGFR1 will be more tolerable than targeting more pleiotropic factors such as AKT²⁹ and presents a rationale for developing more specific FGFR1 antagonists. Regardless, our study provides further evidence that targeting adaptive resistance mechanisms can improve the efficacy of molecular targeted therapies and provides one path towards developing rational strategies for treating KRAS-mutant lung cancer.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 November 2015; accepted 24 May 2016.

Published online 22 June 2016.

1. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nature Rev. Cancer* **3**, 11–22 (2003).
2. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
3. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
4. Scholl, C. et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell* **137**, 821–834 (2009).
5. Luo, J. et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835–848 (2009).
6. Stephen, A. G., Esposito, D., Bagni, R. K. & McCormick, F. Dragging ras back in the ring. *Cancer Cell* **25**, 272–281 (2014).
7. Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: mission possible? *Nature Rev. Drug Discov.* **13**, 828–851 (2014).
8. Lito, P., Rosen, N. & Solit, D. B. Tumor adaptation and resistance to RAF inhibitors. *Nature Med.* **19**, 1401–1409 (2013).
9. Samatar, A. A. & Poulikakos, P. I. Targeting RAS-ERK signalling in cancer: promises and challenges. *Nature Rev. Drug Discov.* **13**, 928–942 (2014).
10. Lito, P. et al. Disruption of CRAF-mediated MEK activation is required for effective MEK inhibition in KRAS mutant tumors. *Cancer Cell* **25**, 697–710 (2014).
11. Sun, C. et al. Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3. *Cell Reports* **7**, 86–93 (2014).
12. Zuber, J. et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nature Biotechnol.* **29**, 79–83 (2011).
13. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
14. Morris, E. J. et al. Discovery of a novel ERK inhibitor with activity in models of acquired resistance to BRAF and MEK inhibitors. *Cancer Discov.* **3**, 742–750 (2013).
15. Lee, H. J. et al. Drug resistance via feedback activation of Stat3 in oncogene-addicted cancer cells. *Cancer Cell* **26**, 207–221 (2014).
16. Nazarian, R. et al. Melanomas acquire resistance to B-Raf(V600E) inhibition by RTK or N-Ras upregulation. *Nature* **468**, 973–977 (2010).
17. Villanueva, J. et al. Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* **18**, 683–695 (2010).
18. Corcoran, R. B. et al. EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2**, 227–235 (2012).
19. Prahallad, A. et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100–103 (2012).
20. Duncan, J. S. et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell* **149**, 307–321 (2012).
21. Kouhara, H. et al. A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell* **89**, 693–702 (1997).
22. Gozgit, J. M. et al. Ponatinib (AP24534), a multitargeted pan-FGFR inhibitor with activity in multiple FGFR-amplified or mutated cancer models. *Mol. Cancer Ther.* **11**, 690–699 (2012).
23. Guagnano, V. et al. Discovery of 3-(2,6-dichloro-3,5-dimethoxy-phenyl)-1-[6-[4-(4-ethyl-piperazin-1-yl)-phenylamino]-pyrimidin-4-yl]-1-methyl-urea (NVP-BGJ398), a potent and selective inhibitor of the fibroblast growth factor receptor family of receptor tyrosine kinase. *J. Med. Chem.* **54**, 7066–7083 (2011).
24. Gavine, P. R. et al. AZD4547: an orally bioavailable, potent, and selective inhibitor of the fibroblast growth factor receptor tyrosine kinase family. *Cancer Res.* **72**, 2045–2056 (2012).
25. Li, F. et al. FGFR-mediated reactivation of MAPK signaling attenuates antitumor effects of imatinib in gastrointestinal stromal tumors. *Cancer Discov.* **5**, 438–451 (2015).
26. Traer, E. et al. Ponatinib overcomes FGF2-mediated resistance in CML patients without kinase domain mutations. *Blood* **123**, 1516–1524 (2014).
27. Jackson, E. L. et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
28. Shao, D. D. et al. KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell* **158**, 171–184 (2014).
29. Shimizu, T. et al. The clinical effect of the dual-targeting strategy involving PI3K/AKT/mTOR and RAS/MEK/ERK pathways in patients with advanced cancer. *Clin. Cancer Res.* **18**, 2316–2325 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sánchez-Céspedes, R. Somwar, and H. Varmus for sharing cell lines; S. Tian, J. Ahn, M. Taylor, A. Shroff, and J. Plevy for technical assistance; C. J. Sherr, L. E. Dow, P. Lito, T. Kasthuber, and J. Leibold for advice on experimental design and/or for editing the manuscript; and other members of the Lowe laboratory for advice and discussions. This work was supported by a program project grant from the National Cancer Institute (S.W.L., N.R.), a grant from the Center of Experimental Therapeutics (S.W.L., N.R.), a Stand Up To Cancer grant from the American Association for Cancer Research (N.R., J.A.E., C.R.), and a Cancer Center Support grant to MSKCC. E.M. was supported by The Jane Coffin Childs Memorial Fund for Medical Research and a K99/R00 grant from the National Institutes of Health/National Cancer Institute. S.W. was supported by the Annette Kade Fellowship from the Watson School of Biological Sciences. R.W. was supported by a Carl-Duisberg Fellowship from the Bayer Foundation. A.L. was supported by an EMBO Long-Term fellowship. E.d.S. received support through the Geoffrey Beene Cancer Research Center. S.W.L. is the Geoffrey Beene Chair of Cancer Biology and a Howard Hughes Medical Institute investigator.

Author Contributions E.M. conceived the project, performed and analysed experiments, and wrote the paper with assistance of all authors. S.W., C.C., and R.W. performed and analysed *in vitro* experiments. S.W., J.P.M., and E.d.S. performed and analysed *in vivo* experiments. A.L. helped design and produce the shRNA library. J.T.P. and C.R. provided and analysed patient-derived xenografts. J.F.G., R.B.C., and J.A.E. provided human specimens. N.R. conceived the project, supervised experiments, and wrote the paper. S.W.L. conceived the project, supervised experiments, analysed data, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.R. (rosenn@mskcc.org) or S.W.L. (lowes@mskcc.org).

Reviewer Information Nature thanks J. Tyner and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Pooled negative-selection RNA interference screening. A custom shRNA library targeting 526 human kinases was designed using miR30-adapted DSIR predictions refined with 'sensor' rules³⁰ (six shRNAs per gene) and constructed by PCR-cloning a pool of oligonucleotides synthesized on customized arrays (Agilent Technologies and CustomArray) as previously described (Supplementary Table 1)¹². The list of genes was obtained from the KinBase Database (<http://kinase.com/human/kinome/>) and was manually curated. After sequence verification, 3,156 shRNAs (5–6 per gene) were combined with 20 positive- and negative-control shRNAs at equal concentrations in one pool. This pool was subcloned into the TRMPV-Neo vector and transduced in triplicates into Tet-on H23 KRAS-mutant lung cancer cells using conditions that predominantly led to a single retroviral integration and represented each shRNA in a calculated number of at least 1,000 cells. Transduced cells were selected for 6 days using 1 mg ml⁻¹ G418 (Invitrogen); at each passage more than 30 million transduced cells were maintained to preserve library representation throughout the experiment. After drug selection, *T*₀ samples were obtained (~30 million cells per replicate (*n* = 3)) and cells were subsequently cultured in the presence or absence of trametinib (25 nM) and 1 µg ml⁻¹ doxycycline to induce shRNA expression. After ten population doublings (*T*_f), about 15 million shRNA-expressing (dsRed⁺/Venus⁺) cells were sorted for each replicate using a FACSariaII (BD Biosciences). Genomic DNA from *T*₀ and *T*_f samples was isolated by two rounds of phenol extraction using Phase Lock tubes (5 Prime) followed by isopropanol precipitation. Deep-sequencing template libraries were generated by PCR amplification of shRNA guide strands as previously described¹². Libraries were analysed on an Illumina Genome Analyzer at a final concentration of 8 pM; 50 nucleotides of the guide strand were sequenced using a custom primer (miR30EcoRISeq, TAGCCCTTGAATTCCGAGGAGTAGGCA). To provide a sufficient baseline for detecting shRNA depletion in experimental samples, we aimed to acquire >500 reads per shRNA in the *T*₀ sample, which required more than 20 million reads per sample to compensate for disparities in shRNA representation inherent to the pooled plasmid preparation or introduced by PCR biases. With these conditions, we acquired *T*₀ baselines of >500 reads for 3,151 (97.9%) shRNAs. Sequence processing was performed using a customized Galaxy platform³¹.

Using selection criteria that required an shRNA depletion averaging greater than fourfold after ten population doublings and an effect greater than fourfold in trametinib-treated cells with respect to untreated ones, 64 shRNAs were identified. The eight targets for which at least two shRNAs were selectively depleted after trametinib treatment were subject to secondary validation in cell competition assays using multiple KRAS-mutant lung cancer cell lines. Six targets validated in the cell line in which the primary screen was performed (H23 cells) and four (*BRAF*, *CRAF*, *ERK2*, and *FGFR1*) across a panel of KRAS-mutant lung cancer cells; as such, these represented the top hits of our screen.

Plasmids and recombinant proteins. All vectors were derived from the Murine Stem Cell Virus (MSCV, Clontech) retroviral vector backbone. miR30- and miRE-based shRNAs were designed and cloned as previously described¹² and sequences are available in Supplementary Table 1. shRNAs were cloned into the TRMPV-Neo (pSIN-TREdsRed-miR30-PGK-Venus-IRES-NeoR), LT3GEPiR (TRE3G-GFP-miRE-PGK-PuroR-IRES-rtTA3), and MLP (LTR-miR30-PGK-PuroR-IRES-GFP) vectors as previously described¹². All constructs were verified by sequencing. Recombinant proteins FGF2 (8910, Cell Signaling), HGF (100-39, Peprotech), EGF (AF-100-15, Peprotech), and NRG1 (100-03, Peprotech) were used at 50 ng ml⁻¹ for 10 min.

Cell culture, compounds, and competitive proliferation assays. H23, H460, H2030, H358, H2122, H820, H3255, and A549 cells were provided by R. Somwar and H. Varmus. H2009, HCT116, SW480, SW620, DLD-1, PaTu 8988t, 3T3, MIAPACA-2, and PANC-1 cells were purchased from the American Type Culture Collection (ATCC). H69, H82, HCC-33, and H446 were provided by C. Rudin. H1975, H1650, Ludlu-1, H1703, PC-14, H2170, SK-MES-1, H520, H522, EBC-1, HCC-15, H441, A-427, and H1299 cells were provided by M. Sanchez-Céspedes. Cell lines were not authenticated. Murine *KRAS*^{G12D}, *p53*^{R270H} cells were derived from a murine lung adenocarcinoma. Cells were maintained in a humidified incubator at 37°C with 5% CO₂, grown in RPMI 1640 or DMEM supplemented with 10% FBS and 100 IU ml⁻¹ penicillin/streptomycin. All cell lines used were negative for mycoplasma.

Trametinib (S2673), SCH7272984 (S7101), Gefitinib (S1025), Crizotinib (S1068), CP-724714 (S1167), Afatinib (S1011), BGJ398 (S2183), AZD4547 (S2801), and Ponatinib (S1490) were obtained from Selleckchem. Drugs for *in vitro* studies were dissolved in DMSO to yield 5 or 10 mM stock solutions and stored at -80°C.

For shRNA experiments, when necessary, human cancer cells were modified to express the ecotropic receptor and rtTA3 by retroviral transduction of MSCV-RIEP (MSCV-rtTA3-IRES-EcoR-PGK-Puro) followed by drug selection (1 µg ml⁻¹

puromycin for 1 week). Cell lines were transduced with ecotropically packaged TRMPV-Neo-shRNA retroviruses or, alternatively, with amphotropically packaged LT3GEPiR-Puro-shRNA lentiviruses, selected with 1 mg ml⁻¹ G418 or 1 µg ml⁻¹ puromycin for 1 week, and treated with 1 µg ml⁻¹ doxycycline to induce shRNA expression.

For competitive proliferation assays, shRNA-transduced cells were mixed with non-transduced cells (8:2) and cultured with doxycycline in the presence or absence of trametinib (25 nM). The relative percentage of Venus⁺/dsRed⁺ or GFP⁺ cells was determined before (*T*₀, blue bars) and after ten population doublings (*T*_f) (results are relative to *T*₀) (*T*_f on doxycycline, grey bars; *T*_f on dox + trametinib, red bars). The quantification of fluorescent cells was monitored on a Guava EasyCyte (Millipore). Experiments were performed independently two or three times.

Lentiviral production. Lentiviruses were produced by co-transfection of 293T cells with lentiviral-Cre backbone construct and packaging and envelope vectors (psPAX2 and VSV-G), using the calcium phosphate method. Supernatant was collected 48, 60, and 72 h after transfection, concentrated by ultracentrifugation at 24,000 r.p.m. for 120 min and resuspended in an appropriate volume of HBSS solution (Gibco).

Clonogenic and apoptosis assay. For clonogenic assays, cells were seeded in triplicate into six-well plates (5 × 10³ to 10 × 10³ cells per well) and allowed to adhere overnight in regular growth media. Cells were then cultured in the absence or presence of drug as indicated in complete media for 10–14 days. Growth media with or without drug was replaced every 2 days. Remaining cells were fixed with methanol (1%) and formaldehyde (1%), stained with 0.5% crystal violet, and photographed using a digital scanner. Relative growth was quantified by densitometry after extracting crystal violet from the stained cells using 10% of acetic acid. All experiments were performed at least three times. Representative experiments are shown.

For apoptosis assays, around 1 × 10⁶ cells were seeded into 10-cm plates and cultured in the presence or absence of drugs as indicated. After 6 days, apoptosis and cell death were determined using AnnexinV-APC apoptosis detection kit according to the manufacturer's instructions (Affymetrix eBioscience). Data were acquired using a FACS Calibur (BD Biosciences). All experiments were performed independently three times.

Quantitative analysis of drug synergy and determination of fold change in sensitivity to trametinib. Drug synergism was analysed using CompuSyn software (version 1.0) (<http://www.combosyn.com>), which is based on the median-effect principle (Chou) and the combination index-isobologram theorem (Chou-Talalay)³². CompuSyn software generates combination index (CI) values, where CI < 0.75 indicates synergism, CI = 0.75–1.25 indicates additive effects, and CI > 1.25 indicates antagonism. Following the instruction of the software, drug combinations at non-constant ratios were used to calculate the combination index in our study.

For calculating the fold change in sensitivity to trametinib, the concentration of trametinib that inhibited cell proliferation by 50% (GI₅₀) was determined for a panel of KRAS wild-type and mutant cancer cell lines in the absence or presence of ponatinib and AZD4547. Experiments were performed independently twice.

Immunoblotting and RAS-GTP assay. Phospho-lysis buffer (50 mM Tris pH 7.5, 1% Tween-20, 200 mM NaCl, 0.2% NP-40) supplemented with phosphatase inhibitors (5 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium pyrophosphate, 1 mM β-glycerophosphate), and protease inhibitors (Protease Inhibitor Cocktail Tablets, Roche) was used for cell lysis, and protein concentration was determined by a Bradford protein Assay kit (Biorad). Proteins were separated by SDS-PAGE and immunoblotted and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore) according to standard protocols. Membranes were immunoblotted with antibodies against pERK^{T202/Y204} (9101), tERK (9107), pAKT^{S473} (4060), tAKT (9272), pFRS2^{Y436} (3861), pSTAT3^{T705} (9145), pMEK^{S217/221} (9154), MEK (4694), pMET^{Y1234/1235} (3077), MET (8198), pERBB2^{Y1221/1222} (2243), pEGFR^{Y1068} (3777), EGFR (4267), pERBB3^{Y1289} (4791), and PTEN (9559) from Cell Signaling; CRAF (SC-227) and BRAF (SC-5284) from Santa Cruz Biotechnology; and KRAS (WH0003845M1) from Sigma in 5% BSA in TBS blocking buffer. After primary antibody incubation, membranes were probed with ECL anti-rabbit IgG, anti-mouse IgG or anti-goat IgG secondary antibody (1:10,000) from GE Healthcare Life Science and imaged using a FluorChem M system (protein simple). GTP-bound RAS was measured using a CRAF RAS-binding-domain (RBD) pull down and detection kit (8821, Cell Signaling) as instructed by the manufacturer. All immunoblots were performed independently at least twice.

qRT-PCR. Total RNA was isolated using TRIzol (Invitrogen), and complementary DNA was obtained using the TaqMan reverse transcription reagents (Applied Biosystems). Real-time PCR was performed in triplicate in three independent experiments using SYBR Green PCR Master Mix (Applied Biosystems) on the

ViiA 7 Real-Time PCR System (Invitrogen). GAPDH or β -actin served as endogenous normalization controls.

Animal studies. All mouse experiments were approved by the Memorial Sloan Kettering Cancer Center (MSKCC) Animal Care and Use Committee (protocol number 12-04-006). Mice were maintained under specific pathogen-free conditions, and food and water were provided *ad libitum*. Female 5- to 7-week-old athymic NCR-NU-NU (Harlan Laboratories) mice were used for animal experiments with human cell lines and patient-derived xenografts. For A549, H23, and H2122 xenografts, cells (10×10^6) were harvested on the day of use and injected in growth-factor-reduced Matrigel/PBS (50% final concentration). One flank was injected subcutaneously per mouse. For JHU-LX55a patient-derived xenograft, a poorly differentiated lung adenocarcinoma bearing a KRAS^{G12C} mutation, tumours were cut into pieces and inserted into a pocket in the subcutaneous space as previously described³³. After inoculation, mice were monitored daily, weighed twice weekly, and calliper measurements begun when tumours became visible. Tumour volume was calculated using the following formula: tumour volume = $(D \times d^2)/2$, in which D and d refer to the long and short tumour diameter, respectively. When tumours reached a size of 150–300 mm³, mice were randomized into five to eight per group and treated with vehicle, trametinib, and/or ponatinib *per os* for 4 consecutive days followed by 3 days off treatment, at 3 mg/kg body weight and 30 mg/kg body weight, respectively. No obvious toxicities were observed in the vehicle- or drug-treated animals as assessed by difference in body weight between vehicle- and drug-treated mice taking tumour size into account. For immunohistochemistry analysis of JHU-LX55a patient-derived xenograft tumours, tumours were harvested 4 h after dosing on day 18.

For drug efficacy studies using a GEMM of lung cancer, *Kras*^{LSL-G12D/+} and *Trp53*^{fl/fl} mice (8–12 weeks old) were anaesthetized by intraperitoneal injection of ketamine (80 mg per kg body weight) and xylazine (10 mg per kg body weight) and infected intratracheally with 2.5×10^5 infectious particles of Lenti-Cre per mouse, as previously described³⁴. Mice were evaluated by micro-computed tomography imaging to quantify lung tumour burden before being assigned to various treatment study cohorts. Mice were treated with vehicle, trametinib and/or ponatinib *per os* for 4 consecutive days followed by 3 days off treatment, at 3 mg/kg body weight and 30 mg/kg body weight, respectively. Micro-computed tomography imaging evaluation was repeated every week during the treatment. Investigators were not blinded with respect to treatment.

For drug efficacy studies using an organoid-derived murine model of pancreatic cancer, spherical, duct-like organoids were derived and cultured in Matrigel and defined media as previously described³⁵ from pancreatic ductal adenocarcinoma (PDAC) occurring in *Kras*^{LSL-G12D/+}, *Trp53*^{fl/+}; *CHC* (untargeted collagen homing cassette); *RIK* (Rosa26-LSL-rtTa3-IRES-Kate2); *p48Cre* mice (GEMM-KPC^{fl/+}) generated via the PDAC–GEMM–ESC approach³⁶. After initially establishing primary organoid cultures, Kate-positive cells were sorted and expanded to minimize injection of non-recombined, normal duct cells. For the orthotopic transplantation of PDAC organoids, mice were anaesthetized using isoflurane, and the pancreas was externalized through a small incision made in the left abdominal side near the spleen. Organoids (approximately 250,000–500,000 cells per mouse) were removed from Matrigel and separated into single cells by trypsinization, washed, and finally resuspended in 25 μ l of Matrigel (BD) diluted 1:1 with cold PBS. The organoid suspension was injected into the tail region of the pancreas using 28-gauge surgical syringes (Hamilton). Successful injection was verified by the appearance of a fluid bubble without signs of intraperitoneal leakage. The abdominal wall was sutured with absorbable Vicryl suture (Ethicon), and the skin was closed with

wound clips (CellPoint Scientific). Mice were evaluated by ultrasound (Vevo 2100, VisualSonics) to quantify pancreas tumour burden before being randomized to various treatment study cohorts. All the treatment mice had similar initial tumour burden. Mice were treated as described above for drug efficacy studies using a GEMM of lung cancer. Investigators were not blinded with respect to treatment.

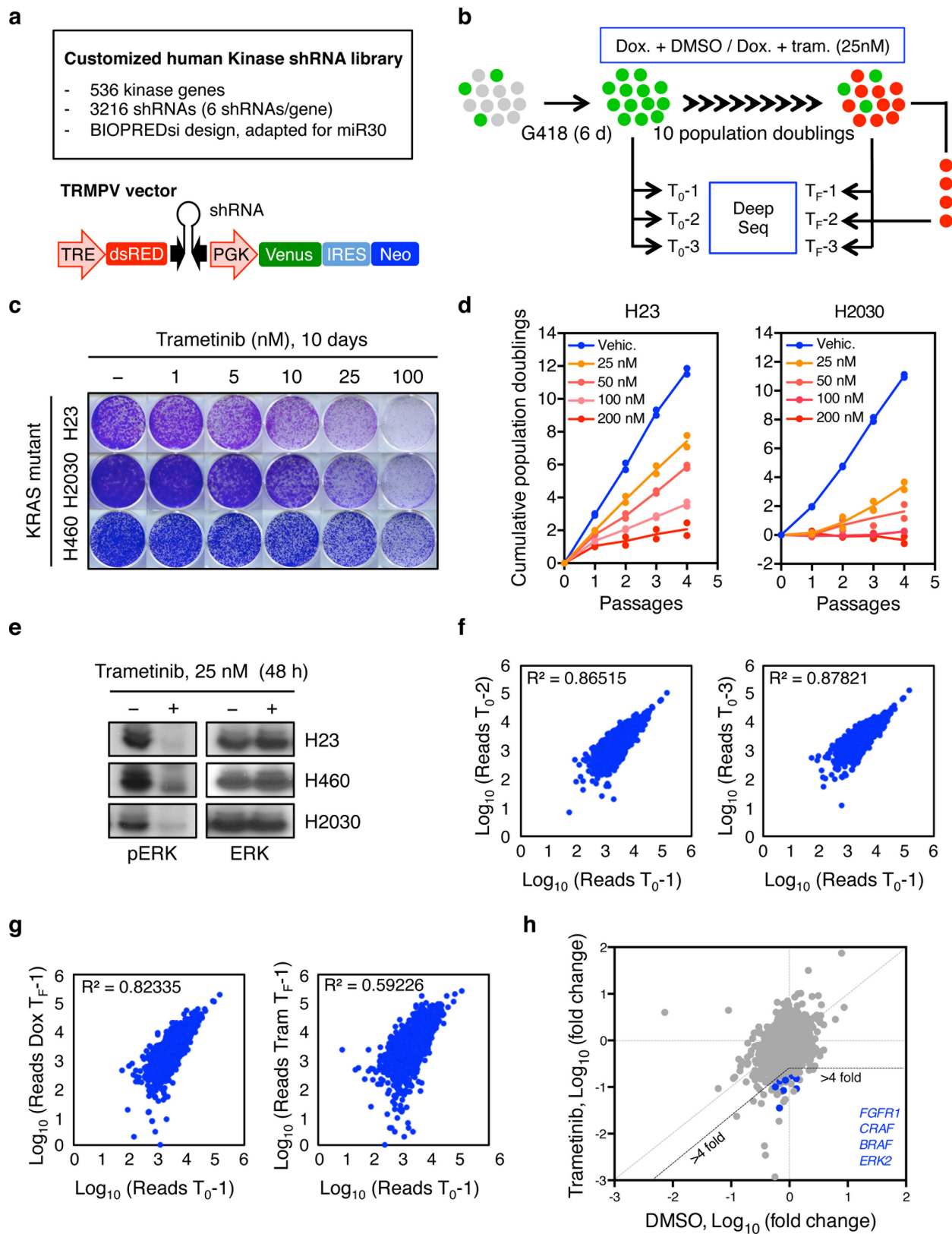
Micro-computed tomography imaging. Micro-computed tomography scans were performed on a Mediso Nano SPECT/CT System covering only the lung fields of each mouse. Each scan averaged approximately 6 min using 240 projections with an exposure time of 1,000 ms set at a pitch of 1°. The tube energy of the X-ray was 55 kVp and 145 μ A. The in-plane voxel sizes chosen were small and thin creating a voxel size of 73 μ m \times 73 μ m \times 73 μ m. The final reconstructed image consisted of 368 voxels \times 368 voxels \times 1,897 voxels. Scans were analysed with Osiris software.

Patients' samples. Patients with KRAS mutation-positive advanced lung adenocarcinomas were enrolled in the phase I/II clinical study of trametinib and navitoclax (NCT02079740) and the response was assessed per RECIST (response evaluation criteria in solid tumours) criteria. Biopsies were obtained before treatment, and within 2–4 weeks after starting the treatment with trametinib. Specifically, for patient 1, the post-treatment biopsy was obtained after treatment with navitoclax for 7 days, followed by co-treatment with navitoclax and trametinib for 16 days. The post-treatment biopsy from patient 2 was obtained after co-treatment with navitoclax and trametinib for 22 days. All human studies were approved by the Massachusetts General Hospital Institutional Review Board, and informed consent to study was obtained as per protocol from all patients.

Immunohistochemistry. Tissues were fixed overnight in 4% paraformaldehyde, embedded in paraffin, and cut into 5 μ m sections. Sections were subjected to haematoxylin and eosin staining, and immunohistochemical staining following standard protocols. The following primary antibodies were used: pERK^{T202/Y204} (4370) and pAKT^{S473} (4060) (Cell signaling), and pFRS2^{Y436} (ab193363) (Abcam).

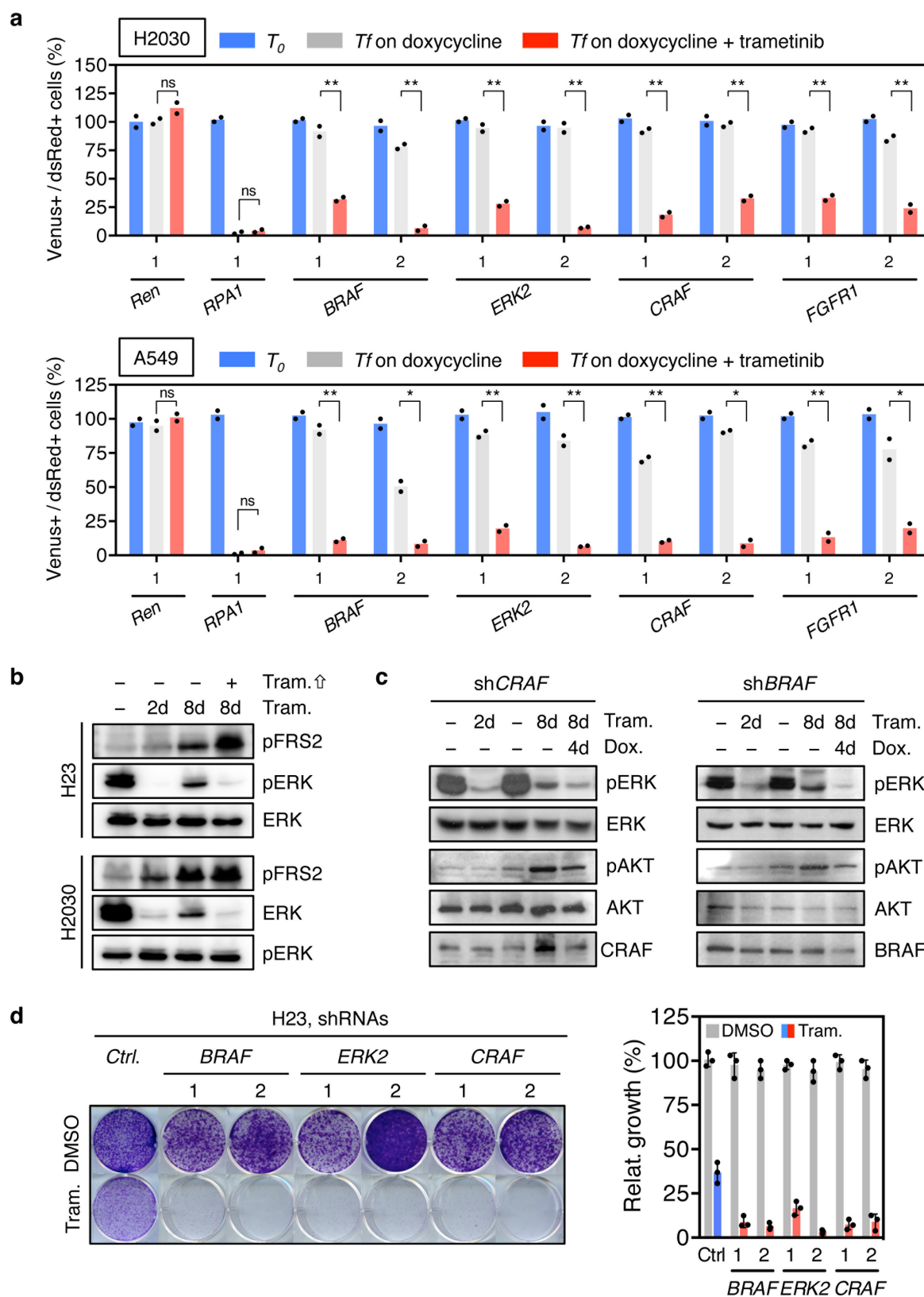
Statistical analysis. Data are expressed as mean \pm s.e.m. or mean \pm s.d. Group size was determined on the basis of the results of preliminary experiments and no statistical method was used to predetermine sample size. The indicated sample size (n) represents biological replicates. Group allocation and outcome assessment were not performed in a blinded manner. All samples that met proper experimental conditions were included in the analysis. Survival was measured using the Kaplan–Meier method. Statistical significance was determined by Student's t -test, log-rank test, and Pearson's correlation using Prism 6 software (GraphPad Software). Significance was set at $P < 0.05$.

30. Fellmann, C. *et al.* Functional identification of optimized RNAi triggers using a massively parallel sensor assay. *Mol. Cell* **41**, 733–746 (2011).
31. Taylor, J., Schenck, I., Blankenberg, D. & Nekrutenko, A. in *Current Protocols in Bioinformatics* (eds Baxevanis, A. D. *et al.*) Ch. 10, Unit 10.5 (Wiley, 2007).
32. Chou, T. C. Drug combination studies and their synergy quantification using the Chou–Talalay method. *Cancer Res.* **70**, 440–446 (2010).
33. Poirier, J. T. *et al.* DNA methylation in small cell lung cancer defines distinct disease subtypes and correlates with high expression of EZH2. *Oncogene* **34**, 5869–5878 (2015).
34. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nature Protocols* **4**, 1064–1072 (2009).
35. Huch, M. *et al.* Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* **32**, 2708–2721 (2013).
36. Saborowski, M. *et al.* A modular and flexible ESC-based mouse model of pancreatic cancer. *Genes Dev.* **28**, 85–97 (2014).



Extended Data Figure 1 | A synthetic lethal RNA interference screen identifies different MAPK signalling effectors and FGFR1 as sensitizers to MEK inhibition in KRAS-mutant lung cancer cells. a, Library features and schematic of the TRMPV-Neo vector. **b,** Schematic outline of the synthetic lethal RNA interference screen for identifying sensitizers to trametinib in KRAS-mutant lung cancer cells. **c,** Clonogenic assay of KRAS-mutant lung cancer cell lines (H23, H460, and H2030) cultured in the presence of increasing concentrations of trametinib. **d,** Proliferation assay of H23 and H2030 cells in the presence of increasing concentrations of trametinib for four passages. Data presented as mean of two independent

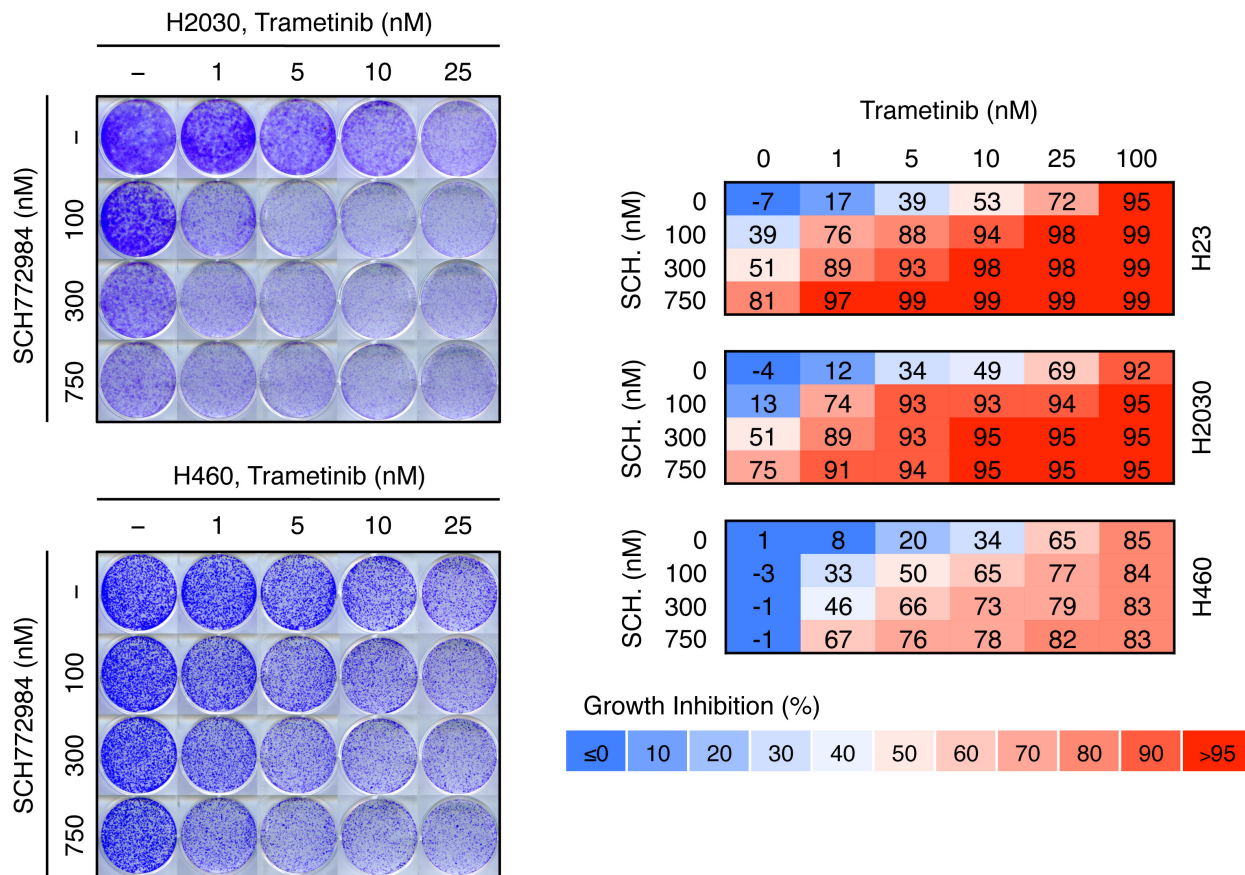
replicates. **e,** Immunoblot analysis of KRAS-mutant lung cancer cell lines treated with 25 nM of trametinib for 48 h. **f, g,** Scatter plots illustrating the correlation of normalized reads per shRNA between replicates at the beginning of the experiment (**d**) and replicates at different time points in the absence (left) or presence (right) of trametinib (25 nM) (**e**). **h,** Scatter plot illustrating the fold change in the relative abundance of each shRNA in the library after ten population doublings on doxycycline in the absence or presence of trametinib (25 nM) in H23 cells. Two shRNAs for *FGFR1*, *CRAF*, *BRAF*, and *ERK2* were identified as selectively depleted in trametinib-treated cells. For gel source data, see Supplementary Fig. 1.



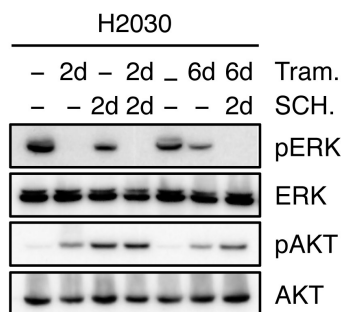
Extended Data Figure 2 | Suppression of FGFR1 and different MAPK signalling effectors reduces the proliferation and viability of KRAS-mutant lung cancer cells treated with trametinib. a, Quantification of fluorescent cells in competitive proliferation assays in H2030 (upper) and A549 (lower) cells transduced with non-targeting control (*Ren*) or the indicated shRNAs. Data presented as mean ($n=2$). Unpaired two-tailed *t*-test. $*P < 0.05$, $**P < 0.01$. **b,** Immunoblot of H23 and H2030 cells pre-treated with 25 nM trametinib for various times and subsequently treated with 200 nM trametinib for 2 h. **c,** Immunoblot of H23 cells transduced

with doxycycline-inducible shRNAs targeting *CRAF* and *BRAF* and treated with trametinib (25 nM) and doxycycline for the times shown. H23 cells were pre-treated with trametinib for 4 days, followed by treatment with doxycycline and trametinib for 4 days. **d,** Clonogenic assay of H23 cells transduced with *BRAF*, *CRAF*, *ERK2*, and non-targeting control shRNAs, and cultured with DMSO or trametinib (25 nM) for 10 days. Relative growth of DMSO- (grey bars) and trametinib-treated cells (blue and red bars) is shown (right). Data presented as mean \pm s.d. ($n=3$). For gel source data, see Supplementary Fig. 1.

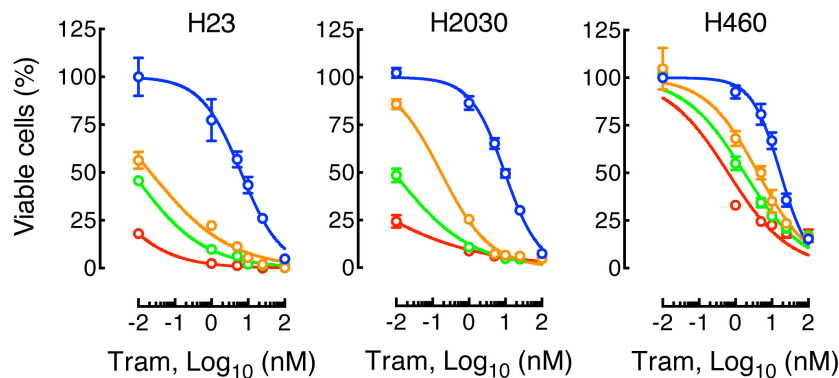
a



b



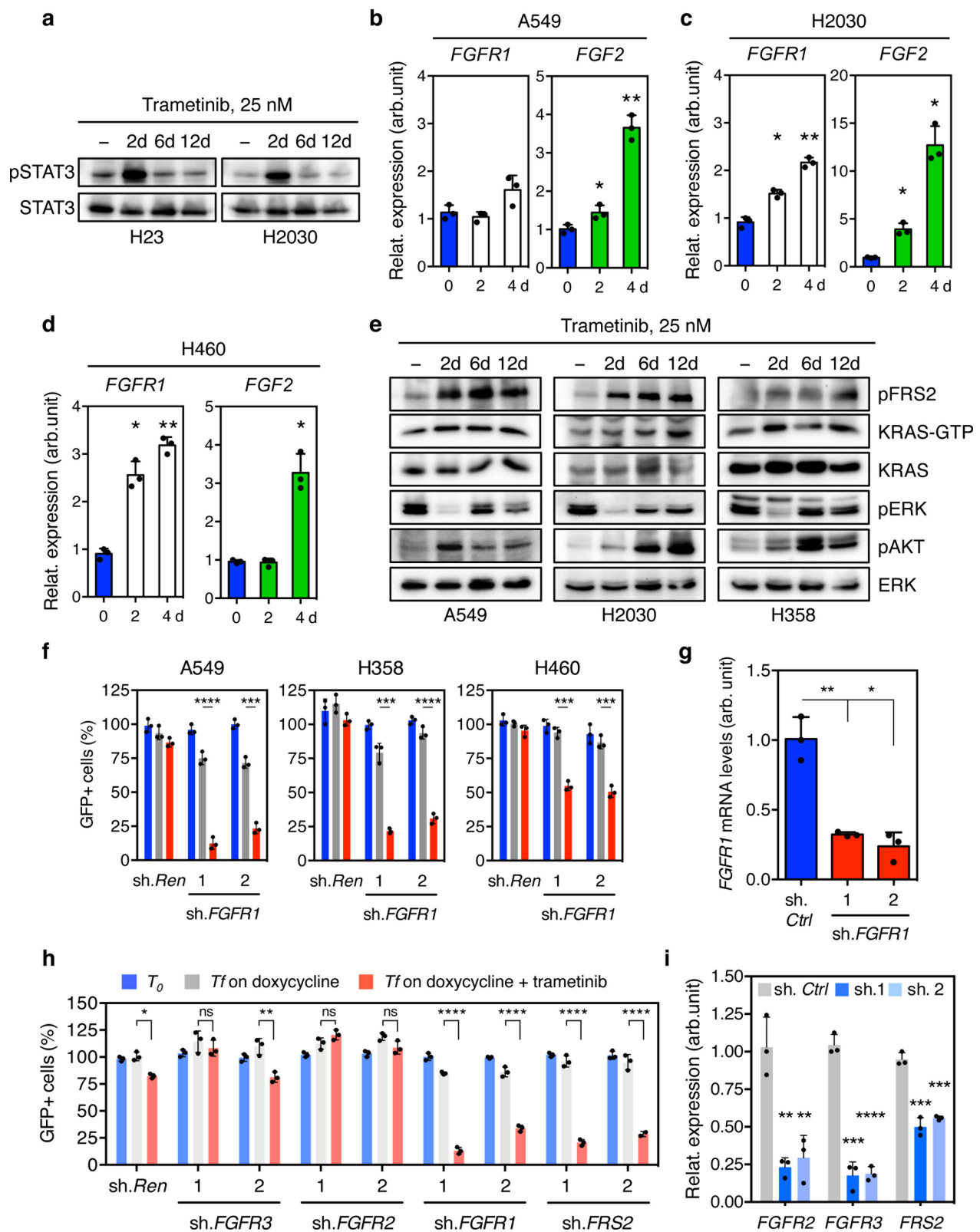
c



		Trametinib, GI ₅₀ (nM)		
		H23	H2030	H460
Ctrl.	●	6.44	9.54	17.16
SCH. 100 nM	●	0.02	0.17	4.18
SCH. 300 nM	●	0.006	0.008	1.75
SCH. 750 nM	●	0.0004	9e-005	0.63

Extended Data Figure 3 | ERK inhibitor SCH772984 enhances the antiproliferative effects of trametinib in KRAS-mutant lung cancer cells. **a**, Clonogenic assay of H2030 (upper) and H460 (lower) cells treated with increasing concentrations of trametinib, ERK inhibitor SCH772984, or their combination as indicated. Percentage inhibition at each concentration of the drugs in H23, H2030, and H460 cells is presented (right). Data presented as mean of three independent experiments ($n = 3$). **b**, Immunoblot analysis of H2030 cells treated with trametinib (25 nM), SCH772984 (500 nM), or their combination for the times shown. H2030

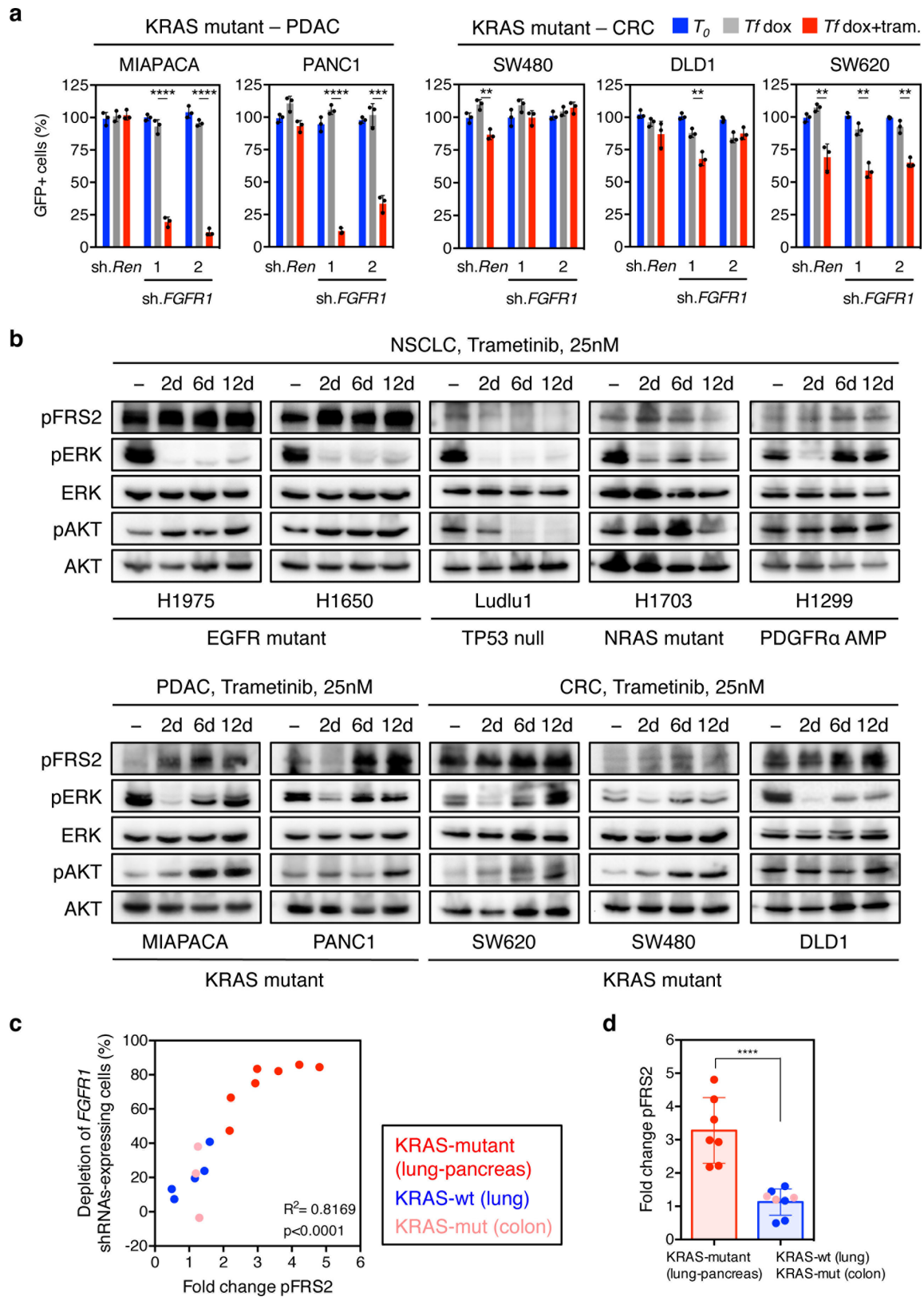
cells were pre-treated with trametinib for 4 days, followed by treatment with SCH772984 and trametinib for 2 days. **c**, Cell viability of H23, H2030, and H460 cells treated with increasing doses of trametinib, ERK inhibitor SCH772984, or their combination for 10 days. Data presented as mean \pm s.d. ($n = 3$). The GI₅₀ was calculated in the absence or the presence of increasing concentrations of SCH772984 (bottom). For gel source data, see Supplementary Fig. 1. Source Data for Extended Data Fig. 3 are available in the online version of the paper.



Extended Data Figure 4 | Feedback activation of FGFR1 signalling leads to adaptive resistance to trametinib in KRAS-mutant lung cancer cells.

a, Immunoblot analysis of KRAS-mutant lung cancer cell lines H23 and H2030 treated with 25 nM trametinib for various times. **b–d**, qRT-PCR for *FGFR1* and *FGF2* in A549 (**b**), H2030 (**c**), and H460 (**d**) cells treated with trametinib for the indicated times. Data presented as mean normalized for *FGFR1* and *FGF2* expression \pm s.d. (n = 3). **e**, Immunoblot analysis of A549, H2030, and H358 cells treated with trametinib (25 nM) for various times. **f**, Quantification of fluorescent cells in competitive proliferation assays in A549, H358, and H460 cells transduced with doxycycline-inducible non-targeting control (*Ren*) or *FGFR1* shRNAs. Data presented

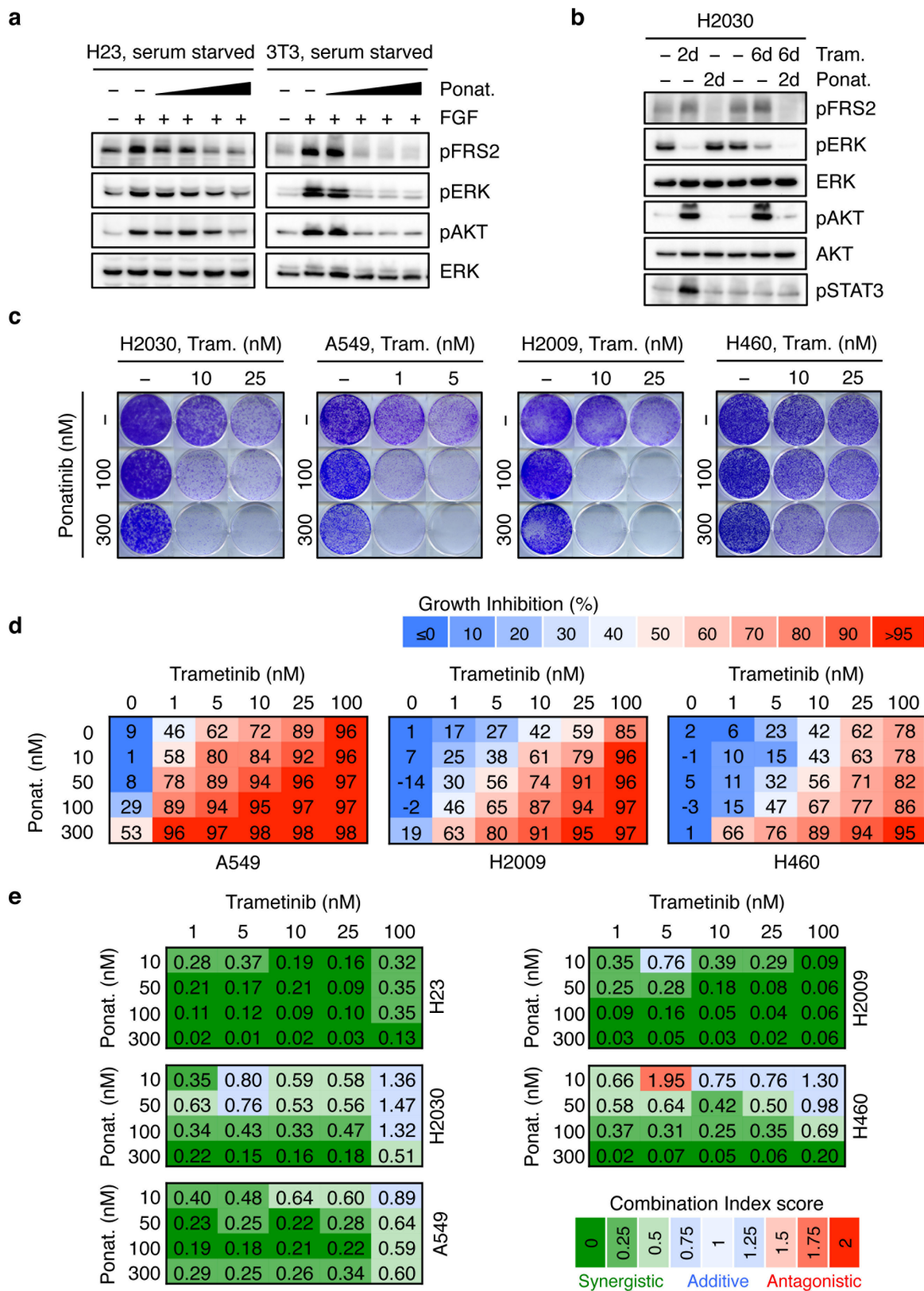
as mean \pm s.d. (n = 3). **g**, qRT-PCR for *FGFR1* in H23 cells transduced with non-targeting control and *FGFR1* shRNAs. Data presented as mean normalized for *FGFR1* expression \pm s.d. (n = 3). **h**, Quantification of fluorescent cells in competitive proliferation assays in A549 cells transduced with non-targeting control (*Ren*) or the indicated shRNAs. Data presented as mean \pm s.d. (n = 3). **i**, qRT-PCR for *FGFR2*, *FGFR3*, and *FRS2* in A549 cells transduced with non-targeting control, *FGFR2*, *FGFR3*, and *FRS2* shRNAs. Data presented as mean normalized for *FGFR2*, *FGFR3*, and *FRS2* expression \pm s.d. (n = 3). **b–d**, Paired two-tailed *t*-test. **f–i**, Unpaired two-tailed *t*-test. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 5 | Trametinib-induced phosphorylation of FRS2 predicts sensitivity to MEK and FGFR1 combined inhibition.

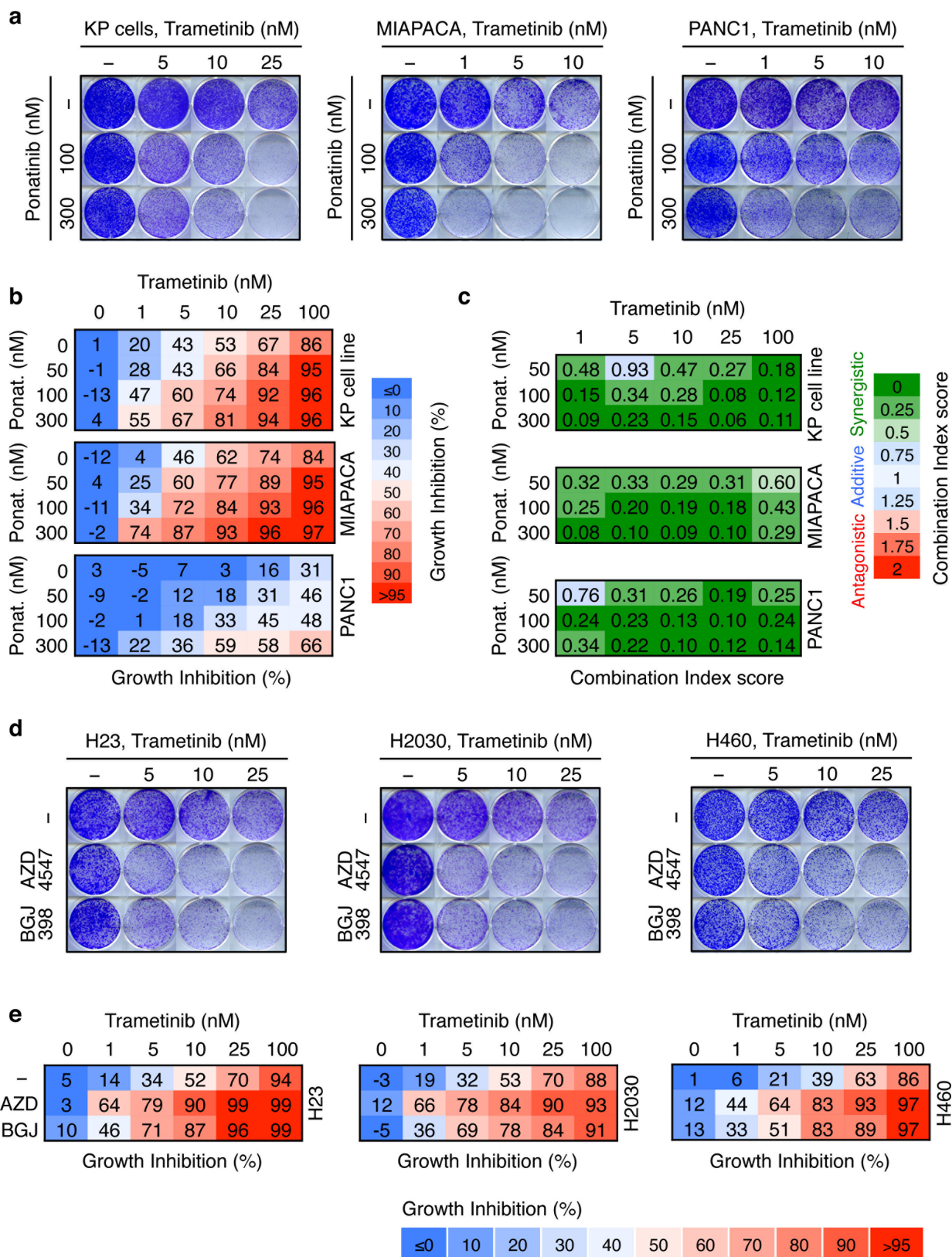
a, Competitive proliferation assays in the indicated KRAS-mutant cancer cell lines transduced with doxycycline-inducible non-targeting control (*Ren*) or *FGFR1* shRNAs. Data presented as mean \pm s.d. ($n=3$). **b**, A panel of lung (H1975, H1650, Ludlu-1, H1703, and H1299), pancreas (MIAPACA, PANC1), and colorectal (SW620, SW480, and DLD1) cancer cell lines were treated with 25 nM trametinib for various times. Lysates were subject

to immunoblot analysis with the indicated antibodies. **c**, Scatter plot illustrating the correlation between depletion of *FGFR1* shRNAs-expressing cells and fold change in FRS2 phosphorylation after trametinib treatment in human cancer cell lines ($n=15$). **d**, Representation of the fold change in FRS2 phosphorylation after treatment with trametinib for 12 days in human cancer cell lines ($n=15$). **a**, **d**, Unpaired two-tailed *t*-test. **c**, Two-tailed Pearson's correlation. $**P<0.01$, $***P<0.001$, $****P<0.0001$. For gel source data, see Supplementary Fig. 1.



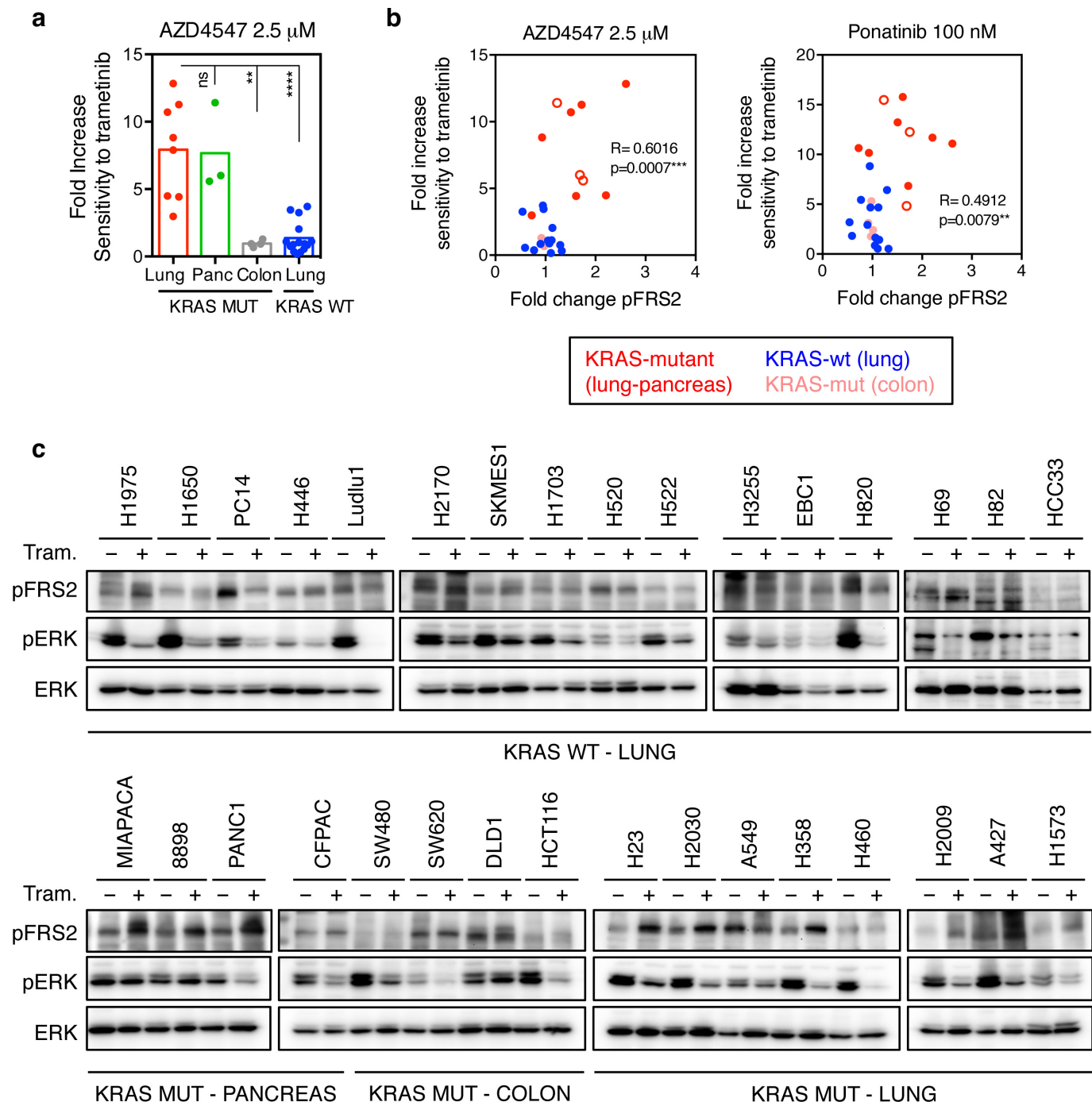
Extended Data Figure 6 | Trametinib in combination with ponatinib synergizes at inhibiting cell proliferation of KRAS-mutant lung cancer cells. **a**, Serum starved H23 (left) and 3T3 (right) cells were pre-treated with increasing concentration of ponatinib for 24 h (1, 30, 100, and 300 nM), followed by stimulation with FGF2 (50 ng ml⁻¹) for 10 min. Immunoblot analysis for the indicated antibodies is shown. **b**, Immunoblot analysis of H2030 cells treated with trametinib (25 nM), ponatinib (750 nM), or their combination for the times shown. Cells were pre-treated with trametinib for 4 days, followed by co-treatment with ponatinib and trametinib for 2 days. **c**, Clonogenic assay of H2030, A549, H2009, and H460 cells treated with

increasing concentrations of trametinib, ponatinib, or their combination as indicated. **d**, Percentage of cell growth inhibition at each concentration of trametinib, ponatinib, or their combination in A549, H2009, and H460 cells after is shown. Data presented as mean of three independent experiments ($n = 3$). **e**, Combination index (CI) scores for H23, H2030, A549, H2009, and H460 cells treated with trametinib in combination with ponatinib at the indicated concentrations. Each CI score represents data from at least three independent experiments. For gel source data, see Supplementary Fig. 1. Source Data for Extended Data Fig. 6 in the online version of the paper.



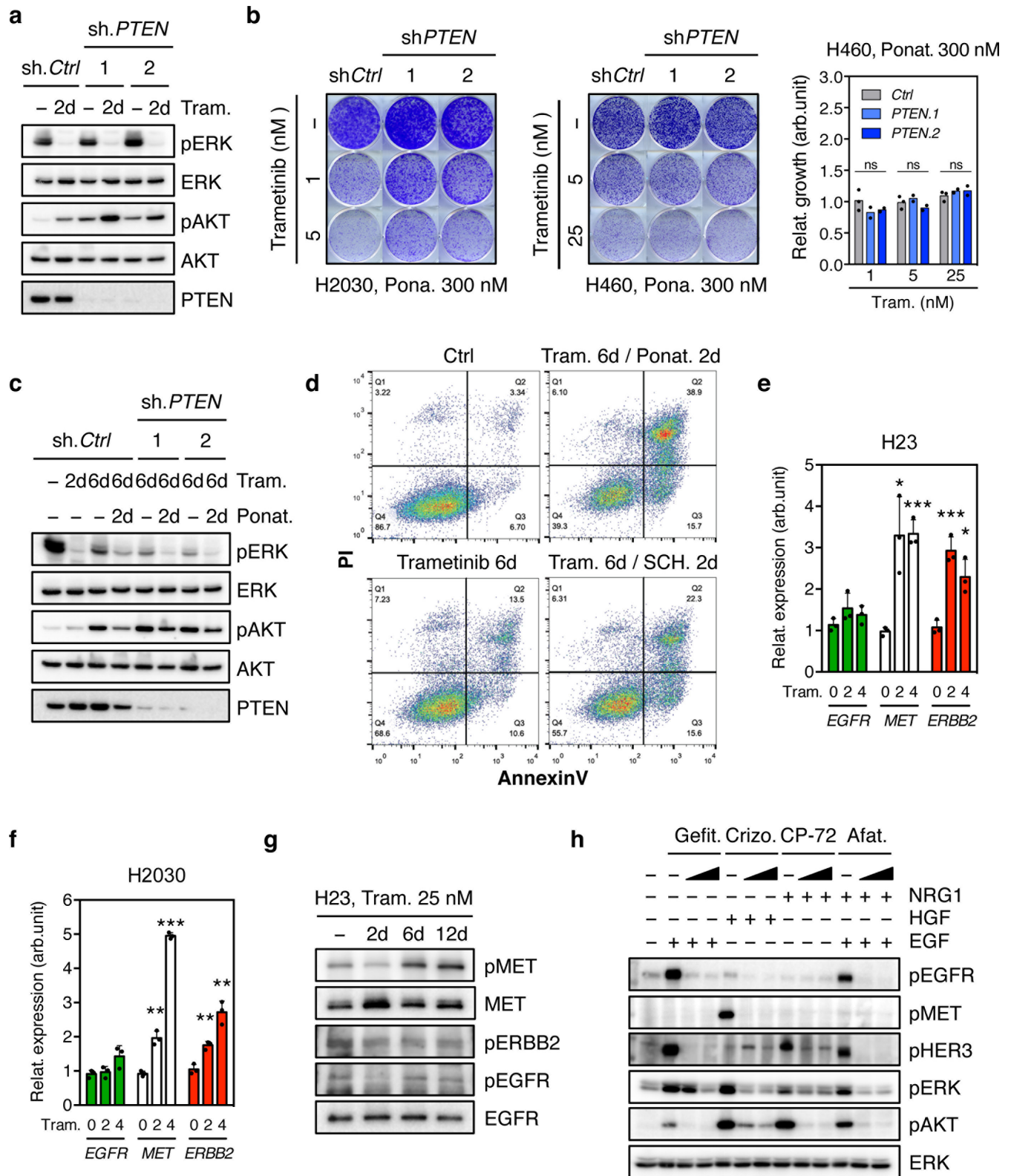
Extended Data Figure 7 | Distinct FGFR1 inhibitors sensitize murine and human KRAS-mutant cancer cells to trametinib. **a**, Clonogenic assay of a murine lung cancer cell line harbouring *Kras*^{G12D} and *Trp53*^{R270H} mutations (KP cell line), and human KRAS-mutant pancreatic cancer cell lines (MIAPACA and PANC1). Tumour cells were cultured with increasing concentrations of trametinib, ponatinib, or their combination as indicated. **b**, Percentage of cell growth inhibition at each concentration of trametinib, ponatinib, or their combination in KP, MIAPACA, and PANC1 cells is shown. Data presented as mean of three independent replicates ($n = 3$). **c**, Combination index scores for KP, MIAPACA, and PANC1 cells

treated with trametinib in combination with ponatinib at the indicated concentrations. Each combination index score represents data from at least three independent experiments ($n = 3$). **d**, Clonogenic assay of H23, H2030, and H460 cells cultured with increasing concentrations of trametinib alone or in combination with FGFR1 inhibitors BGJ398 (1.5 μ M) or AZD4547 (2 μ M). **e**, Percentage of cell growth inhibition at each concentration of trametinib alone or in combination with BGJ398 (1.5 μ M) or AZD4547 (2 μ M) in H23, H2030, and H460 cells is shown. Data presented as mean of three independent replicates ($n = 3$). Source Data for Extended Data Fig. 7 is available in the online version of the paper



Extended Data Figure 8 | The magnitude of trametinib-induced FRS2 phosphorylation correlates with the sensitivity to trametinib and FGFR1 combined inhibition in human cancer cells. **a**, Dot plot illustrating the sensitivity increase to trametinib after the treatment with AZD4547 (2.5 μ M) in a panel of KRAS-mutant ($n = 15$) and KRAS wild-type ($n = 15$) cancer cell lines. Data presented as mean of two independent replicates ($n = 2$). **b**, Scatter plot illustrating the correlation

between fold increase in sensitivity to trametinib after treatment with AZD4547 (2.5 μ M) or ponatinib (100 nM) and fold change in FRS2 phosphorylation after trametinib treatment in a panel of human cancer cells lines. **c**, Immunoblot analysis of a panel of human cancer cells treated with trametinib (25 nM) for 6 days. **a**, Unpaired two-tailed t -test. **b**, Two-tailed Pearson's correlation. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For gel source data, see Supplementary Fig. 1.

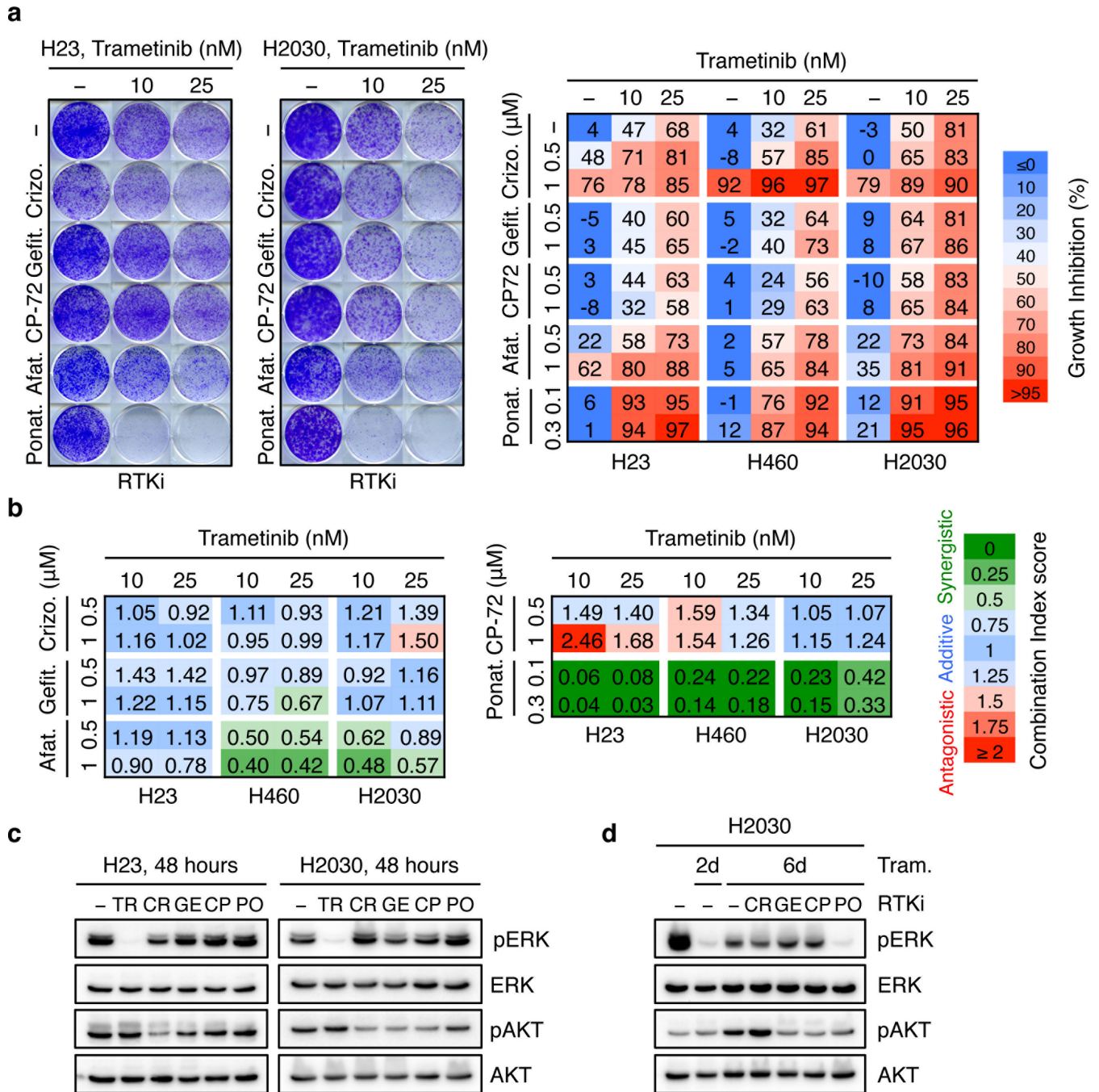


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Ponatinib prevents trametinib-induced reactivation of MAPK and PI3K signalling. Upregulation of distinct RTKs in KRAS-mutant lung cancer cells after trametinib treatment.

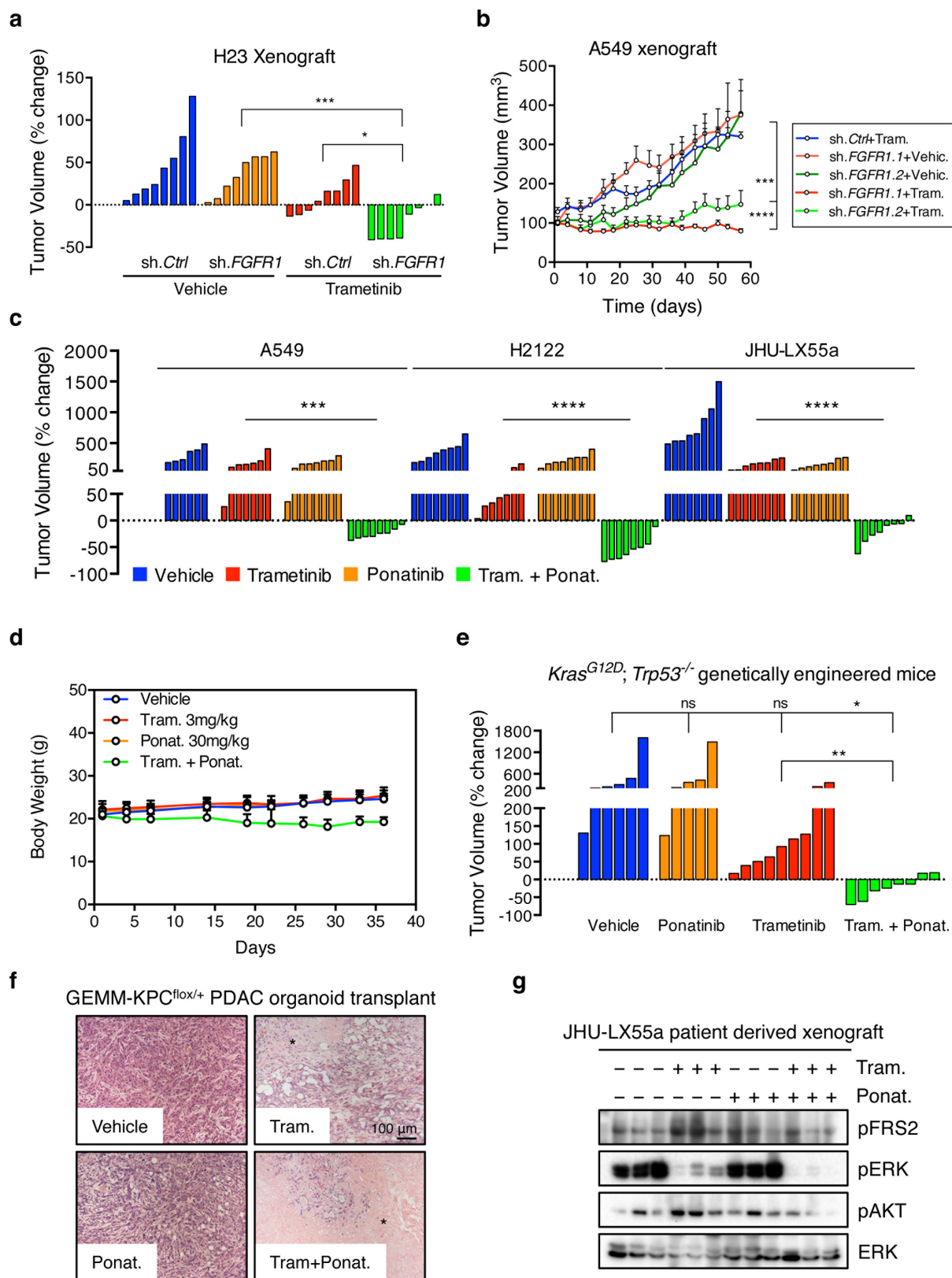
a, Immunoblot analysis of H2030 transduced with *PTEN* and non-targeting control shRNAs, and treated with trametinib (25 nM) for the times shown. **b**, Clonogenic assay of H2030 (left) and H460 (middle) cells transduced with *PTEN* and non-targeting control shRNAs. Cells were treated with ponatinib alone (300 nM) or in combination with trametinib at the indicated concentrations. Quantification of the relative cell growth of H460 cells is shown (right). Data presented as mean of two independent experiments. **c**, Immunoblot analysis of H2030 transduced with *PTEN* and non-targeting control shRNAs, and treated with trametinib (25 nM) alone or in combination with ponatinib (750 nM) for the times shown. *PTEN* suppression did not affect ERK signalling or its inhibition after trametinib

treatment but instead activated AKT and, more importantly, attenuated the ability of ponatinib to suppress trametinib-induced increase in pAKT. **d**, AnnexinV/PI double staining assay of H23 cells treated with vehicle, trametinib (25 nM) alone or in combination with ponatinib (300 nM) or SCH772984 (1 μ M) for the times shown ($n = 3$). **e**, **f**, qRT-PCR for *EGFR*, *MET*, and *ERBB2* in H23 (**e**) and H2030 (**f**) cells treated with trametinib for 0, 2, and 4 days. Data presented as mean normalized for *EGFR*, *MET*, and *ERBB2* expression \pm s.d. ($n = 3$). **g**, Immunoblot analysis of H23 cells treated with 25 nM of trametinib for various times. **h**, Immunoblot analysis of serum starved H2030 cells pre-treated with 500 nM or 1 μ M of gefitinib, crizotinib, CP-724714, or afatinib for 12 h, followed by stimulation with EGF, HGF, NRG1, or their combination (50 ng ml⁻¹) for 10 min. **b**, **e**, **f**, Unpaired two-tailed *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 10 | Unresponsiveness of KRAS-mutant lung cancer cells to MEK inhibitor trametinib is predominantly mediated by feedback activation of FGFR1 signalling. **a**, Clonogenic assay of H23 and H2030 cells treated with increasing concentration of trametinib alone or in combination with 500 nM crizotinib, gefitinib, CP-724714, and afatinib, or 300 nM ponatinib. Percentage inhibition at each concentration of the drugs in H23, H460, and H2030 cells is presented (right). Data presented as mean of at least two independent experiments ($n = 2$). **b**, CI (combination index) scores for H23, H460, and H2030 cells treated with trametinib in combination with crizotinib, gefitinib, CP-724714, afatinib, and ponatinib

at the indicated concentrations. Each CI score represents data from at least two independent experiments ($n = 2$). **c**, Immunoblot of H23 and H2030 treated with trametinib (25 nM), crizotinib (1 μM), gefitinib (1 μM), CP-724714 (1 μM), and ponatinib (750 nM) for 48 h. **d**, Immunoblot analysis of H2030 treated with trametinib (25 nM), crizotinib (1 μM), gefitinib (1 μM), CP-724714 (1 μM), ponatinib (750 nM), or their combination for the times shown. Cells were pre-treated with trametinib for 4 days, followed by co-treatment with RTK inhibitors and trametinib for 2 days. For gel source data, see Supplementary Fig. 1. Source Data for Extended Data Fig. 10 are available in the online version of this paper.



Extended Data Figure 11 | See next page for caption.

Extended Data Figure 11 | Suppression of FGFR1 cooperates with trametinib to inhibit growth of KRAS-mutant lung tumours. **a, b,** Mice bearing H23 (**a**) or H2030 (**b**) xenografts transduced with *FGFR1* or non-targeting control shRNAs were treated with either vehicle or trametinib (3 mg/kg body weight). For H23 xenografts, a waterfall representation of the best response for each tumour is shown ($n = 8$ per group) (**a**). For H2030 xenografts, the tumour volumes are shown as a function of time after treatment. Error bars, mean \pm s.e.m. ($n \geq 4$ per group) (**b**). **c,** Mice bearing A549 and H2122 xenografts, and JHU-LX55a patient-derived xenograft tumours were treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination. A waterfall representation of the best response for each tumour is shown ($n \geq 6$ per group). **d,** Body weight of mice bearing A549 xenografts and treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination for the indicated times ($n \geq 6$ per group). **e,** *Kras*^{G12D}; *Trp53*^{-/-} genetically engineered mice

harbouring lung adenocarcinomas were treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination for 7 weeks. A waterfall representation of the response for each tumour after 7 weeks of treatment is shown ($n \geq 5$). **f,** Representative haematoxylin and eosin stains of pancreatic tumour tissue resulting from orthotopic transplantation of GEMM-KPC^{fllox/+} PDAC organoids. Mice were treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination. Black asterisk indicates necrosis. **g,** Immunoblot analysis of tumour tissue from mice bearing JHU-LX55a patient-derived xenografts treated with vehicle, trametinib (3 mg/kg body weight), ponatinib (30 mg/kg body weight), or both drugs in combination for 18 days. **a–c, e,** Unpaired two-tailed *t*-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For gel source data, see Supplementary Fig. 1. Source Data for Extended Data Fig. 11 are available in the online version of this paper.

The landscape of accessible chromatin in mammalian preimplantation embryos

Jingyi Wu^{1,2*}, Bo Huang^{3*}, He Chen⁴, Qiangzong Yin¹, Yang Liu^{2,5}, Yunlong Xiang¹, Bingjie Zhang¹, Bofeng Liu¹, Qiuju Wang¹, Weikun Xia¹, Wenzhi Li⁶, Yuanyuan Li¹, Jing Ma¹, Xu Peng⁷, Hui Zheng¹, Jia Ming⁶, Wenhao Zhang¹, Jing Zhang⁸, Geng Tian⁹, Feng Xu^{7,10}, Zai Chang⁸, Jie Na⁶, Xuerui Yang^{2,5} & Wei Xie^{1,2}

In mammals, extensive chromatin reorganization is essential for reprogramming terminally committed gametes to a totipotent state during preimplantation development. However, the global chromatin landscape and its dynamics in this period remain unexplored. Here we report a genome-wide map of accessible chromatin in mouse preimplantation embryos using an improved assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq) approach with CRISPR/Cas9-assisted mitochondrial DNA depletion. We show that despite extensive parental asymmetry in DNA methylomes, the chromatin accessibility between the parental genomes is globally comparable after major zygotic genome activation (ZGA). Accessible chromatin in early embryos is widely shaped by transposable elements and overlaps extensively with putative *cis*-regulatory sequences. Unexpectedly, accessible chromatin is also found near the transcription end sites of active genes. By integrating the maps of *cis*-regulatory elements and single-cell transcriptomes, we construct the regulatory network of early development, which helps to identify the key modulators for lineage specification. Finally, we find that the activities of *cis*-regulatory elements and their associated open chromatin diminished before major ZGA. Surprisingly, we observed many loci showing non-canonical, large open chromatin domains over the entire transcribed units in minor ZGA, supporting the presence of an unusually permissive chromatin state. Together, these data reveal a unique spatiotemporal chromatin configuration that accompanies early mammalian development.

The state of chromatin dictates fundamental cellular processes including gene expression, DNA replication and DNA repair¹. It has long been observed that accessible chromatin marks regulatory sequences such as promoters, enhancers, insulators and the locus-control regions². These elements interact with cell-type specific transcription factors to execute transcriptional programs that instruct cell fate determination and development³. Early animal embryos undergo extensive reprogramming and chromatin remodelling to allow the conversion of terminally differentiated gametes to totipotent/pluripotent cells⁴. However, it remains to be determined how mammalian chromatin reconfigures and regulates the transcription programs in preimplantation development. It is also unclear whether the dynamics of chromatin are associated with other epigenomic reprogramming events such as global DNA demethylation⁵. Moreover, the roles of *cis*-regulatory elements in early development are poorly understood. Unfortunately, these questions are difficult to address owing to the limited experimental materials available. Recently, the highly efficient approach ATAC-seq (an assay for transposase-accessible chromatin using sequencing) was developed, which can probe open chromatin using as little as single cells^{6–8}. Accessible regions detected by ATAC-seq colocalize extensively with regulatory elements such as enhancers and promoters^{6,9,10}. Here, we use an improved ATAC-seq approach to interrogate genome-wide accessible regions in mouse preimplantation embryos. As described below, our integrative analyses provide a comprehensive view of the spatiotemporal chromatin configuration that accompanies early embryogenesis.

Mapping accessible chromatin in early embryos

We sought to investigate accessible chromatin in mouse preimplantation embryos using ATAC-seq (Fig. 1a). First, we confirmed that ATAC-seq using 200 and 1,000 mouse embryonic stem cells (mESCs) recapitulated results using 50,000 cells (Fig. 1b and Extended Data Fig. 1a, b). Next, we collected gametes and preimplantation embryos by crossing C57BL/6N female mice with the DBA/2N male mice. We obtained poor ATAC-seq enrichment in sperm, metaphase II oocytes and pronuclear stage 5 zygotes despite multiple attempts (data not shown). For the 2-cell, 4-cell, 8-cell embryos and inner cell masses (ICMs), we carried out ATAC-seq for two biological replicates. One obstacle for profiling accessible chromatin in preimplantation embryos is the abundant mitochondrial DNA (mtDNA) in the ATAC-seq sequencing libraries (Extended Data Fig. 1c). Therefore, we developed a method to deplete mtDNA from the ATAC-seq libraries, named CRISPR/Cas9-assisted removal of mitochondrial DNA (CARM) (Fig. 1a and Extended Data Fig. 1d). Using a pool of single guide RNAs (sgRNAs) targeting mtDNA, CARM substantially increased the percentages of nuclear DNA (Extended Data Fig. 1c, e, f) without affecting ATAC-seq enrichment patterns (Extended Data Fig. 1g). A similar strategy was used to deplete ribosomal RNA-derived sequencing reads in a previous study¹¹. Using ATAC-seq coupled by CARM, we obtained 40–67 million monoclonal reads (replicates combined) uniquely mapped to the nuclear genome for each stage (Supplementary Table 1). We also performed and validated RNA-seq for each stage using Smart-seq¹² (Extended Data Fig. 2a).

¹MOE Key Laboratory of Bioinformatics, Center for Stem Cell Biology and Regenerative Medicine, THU-PKU Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. ²Joint Graduate Program of Peking-Tsinghua-NIBS, School of Life Sciences, Tsinghua University, Beijing 100084, China. ³PKU-THU Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. ⁴Joint Graduate Program of Peking-Tsinghua-NIBS, College of Life Sciences, Peking University, Beijing 100871, China. ⁵MOE Key Laboratory of Bioinformatics, Center for Synthetic & Systems Biology, THU-PKU Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. ⁶Center for Stem Cell Biology and Regenerative Medicine, School of Medicine, Tsinghua University, Beijing 100084, China. ⁷Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore 117609, Singapore. ⁸School of Life Sciences, Tsinghua University, Beijing 100084, China. ⁹School of Medicine, Tsinghua University, Beijing 100084, China. ¹⁰Institute of Molecular and Cell Biology, A*STAR, Singapore 138673, Singapore.

*These authors contributed equally to this work.

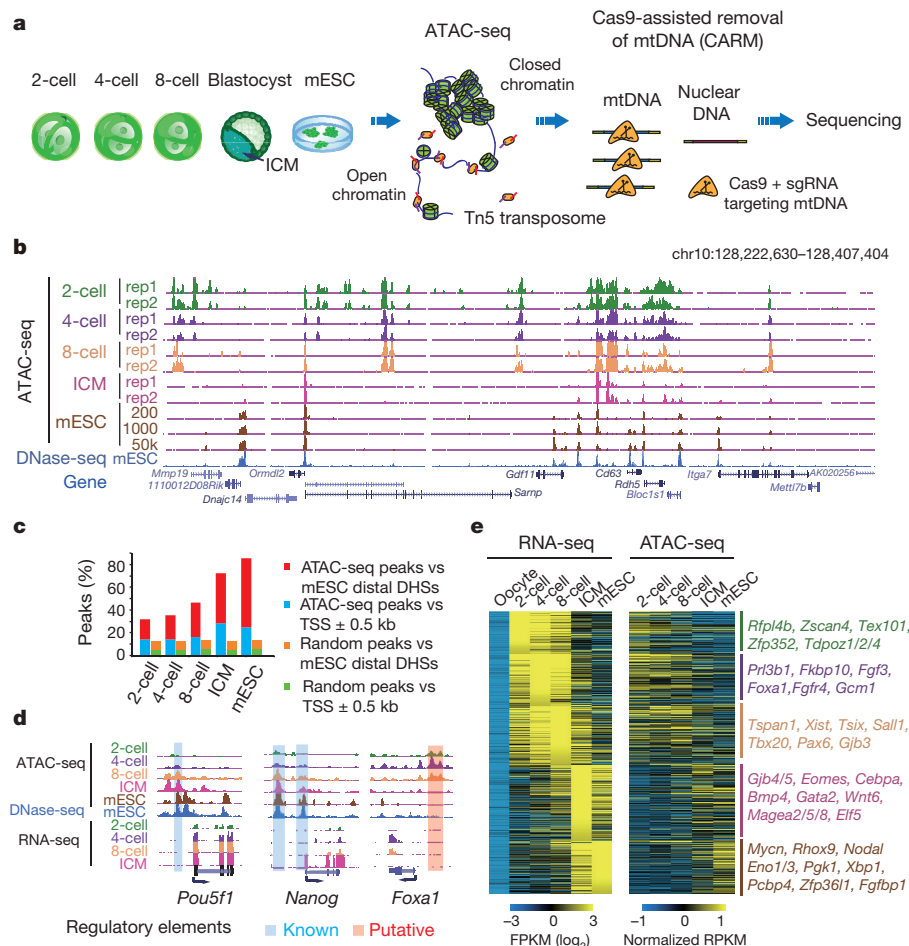


Figure 1 | Accessible chromatin landscape in mouse preimplantation embryos. **a**, Schematic of ATAC-seq and CARM for probing accessible chromatin in mouse preimplantation embryos. **b**, The UCSC browser view shows enrichment of ATAC-seq in early embryos. **c**, The overlap between ATAC-seq peaks and annotated promoters (TSS \pm 0.5 kb) or distal DHSs (away from TSS \pm 0.5 kb) in mESCs. A random set of peaks that match the lengths of individual ATAC-seq peaks on the same chromosomes

were used as a control. **d**, The UCSC browser views showing the ATAC-seq, RNA-seq and DNaseI-seq (mESC, ENCODE) enrichment near *Pou5f1*, *Nanog*, and *Foxa1*. **e**, Heat maps show the expression (FPKM) of stage-specific genes (ZGA-only) (left) and the ATAC-seq enrichment (normalized reads per kilobase per million (RPKM)) of stage-specific genes (ZGA-only) (left) and the ATAC-seq enrichment (normalized reads per kilobase per million (RPKM)) of stage-specific genes (ZGA-only) (left).

We conducted several analyses to validate the ATAC-seq data in early embryos. First, biological replicates of ATAC-seq showed highly reproducible results (Fig. 1b and Extended Data Fig. 1a and 2b). Second, a large fraction of ATAC-seq peaks in early embryos overlapped with promoters and mESC distal DNase I hypersensitive sites (DHSs) (Fig. 1c). Third, for developmentally regulated genes, such as *Pou5f1*, *Nanog* and *Foxa1*, we found elevated ATAC-seq enrichment at annotated or putative enhancers and promoters at the same stages when they are expressed (Fig. 1d). Fourth, the activation of genes at various stages was generally correlated with increased levels of promoter ATAC-seq signals (Fig. 1e). Consistent with previous work¹³, the correlation is more evident for promoters with low CG densities (Extended Data Fig. 2c). As RNA transcripts can also be inherited from oocytes, we only analysed genes activated in preimplantation embryos but not expressed in oocytes (denoted hereafter as 'ZGA-only genes'). Finally, we investigated histone modifications using a low-input chromatin immunoprecipitation (ChIP) followed by high-throughput DNA sequencing (ChIP-seq) method that we recently developed (Methods; manuscript submitted) to validate the ATAC-seq data. In the 2-cell embryos, the ATAC-seq enrichment showed similar distributions with the active mark H3K27ac but not with the repressive mark H3K27me3 (Extended Data Fig. 2d, e). Taken together, these data suggest that ATAC-seq provides a highly sensitive method to examine open chromatin and *cis*-regulatory element activities in preimplantation embryos.

Allelic landscape of open chromatin

To determine the global chromatin state on each parental genome, we measured allelic ATAC-seq enrichment using single-nucleotide polymorphisms (SNPs) present between the two parental strains (Fig. 2a). We interrogated regions covered with sufficient reads of which the allele origins were identified (≥ 8 reads per 1 kb window; Methods). These regions overlapped with 38–39% of the total ATAC-seq peaks at each stage. Among them, a large fraction (82–88%) did not show significant allelic bias for ATAC-seq enrichment (false discovery rate (FDR) $< 2\%$; Fig. 2b and Extended Data Fig. 3a, b). Consistently, we found that the majority of ZGA-only genes are biallelically transcribed in pooled embryos (Extended Data Fig. 3b). Conversely, we also observed allele-specific ATAC-seq signals in early embryos (Extended Data Fig. 3c). To determine whether this reflects differences in parent-of-origins, sequences or stochastic events, we chose one stage (4-cell) and conducted ATAC-seq analysis in the reciprocal cross by mating male C57BL/6N and female DBA/2N mice (with two replicates). Among all ATAC-seq peaks identified as allele-specific in at least one cross, the majority (72%) did not show allele-specific enrichment in the other cross ('stochastic') (Fig. 2c). About 22% showed preferential enrichment on the allele from one strain ('sequence dependent'). Finally, only 6% showed parental-dependent allele-specific open chromatin ('parent-of-origin') (Fig. 2c). Taken together, about 5.3% of all SNP-covered ATAC-seq

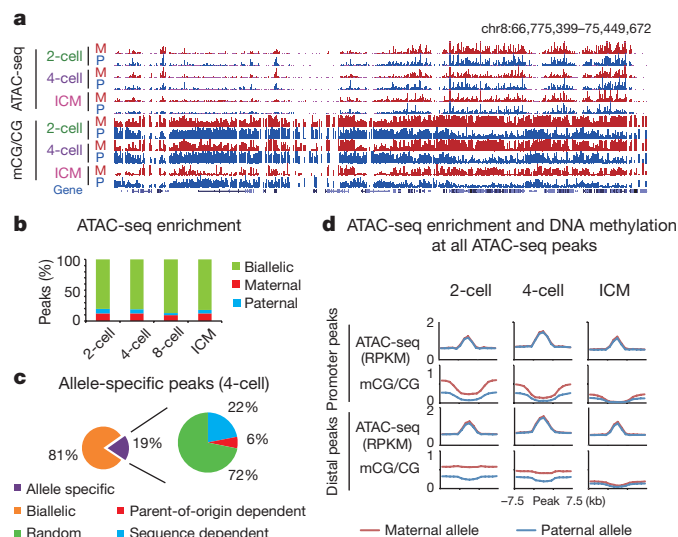


Figure 2 | Allele-specific accessible chromatin landscape in mouse early embryos. **a**, The UCSC browser view shows allelic ATAC-seq enrichment and DNA methylation. mCG/CG data from ref. 16. **b**, Bar charts showing the percentages of ATAC-seq peaks that are biallelic, paternal- or maternal-specific. Only peaks covered by sufficient SNPs and allelic reads were considered (Methods). **c**, Pie charts show the percentages of peaks (covered by SNPs) with biallelic or allele-specific accessible chromatin at the 4-cell stage. **d**, The average ATAC-seq enrichment and DNA methylation levels on each allele around all promoter and distal ATAC-seq peaks (covered by SNPs).

peaks were reproducibly identified as allele-specific peaks (including 1.1% for parent-of-origin-dependent and 4.2% for sequence-dependent; Supplementary Table 2). Regions on the X-chromosomes tended to show lower paternal enrichment (Extended Data Fig. 3d), which is in part contributed to by the lower numbers of paternal chromosomes in the pooled embryo samples (X_p/X_m in females + X_m/Y in males = $2X_m:1X_p$). Other parent-of-origin-dependent allele-specific peaks include those near known imprinted gene *Snrpn* and a novel imprinted gene *Etv6*, a transcription factor that is essential for embryonic development¹⁴ (Extended Data Fig. 3c). Globally, genes with allele-specific expression are preferentially located near allele-specific ATAC-seq peaks (Extended Data Fig. 3e). These data suggest that the two parental genomes showed comparable accessible chromatin and transcription activities after ZGA at the bulk levels, with a small subset of regions showing allele-specific open chromatin and transcription.

Allele-specific expression of imprinted genes is controlled by differential DNA methylation¹⁵. Therefore, we asked if allele-specific ATAC-seq peaks are associated with allelic DNA methylation. A global view showed that the comparable allelic landscape of accessible chromatin is in stark contrast to the distinct allelic DNA methylomes previously reported¹⁶ (Fig. 2a). For both promoter and distal ATAC-seq peaks, DNA methylation shows preferential methylation on the maternal allele (Fig. 2d). This is true for biallelic, maternal-specific, and paternal-specific peaks (Extended Data Fig. 4a) and still holds when considering the demethylation products such as 5-formylcytosine and 5-hydroxymethylcytosine¹⁶ (Extended Data Fig. 4b). These results are consistent with the higher level of global DNA methylation on the maternal genome at the 2-cell stage¹⁶. The differences in allelic DNA methylation at ATAC-seq peaks gradually decrease when development proceeds, as both alleles are progressively demethylated¹⁶ (Fig. 2d). By contrast, imprinted regions maintain allele-specific DNA methylation throughout early development¹⁶ (Extended Data Fig. 4c). These data indicate that in early embryos, maternal and paternal genomes showed largely comparable open chromatin landscapes and zygotic transcriptomes despite the presence of widespread allele-specific DNA methylation.

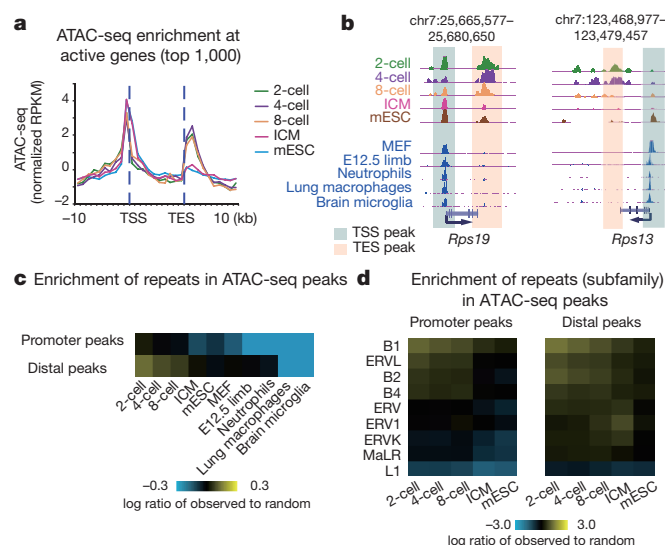


Figure 3 | Unique characteristics of accessible chromatin in mouse early embryos. **a**, The average ATAC-seq enrichment (normalized) for the top 1,000 most active genes (ZGA-only genes) at each stage. **b**, The UCSC browser view shows ATAC-seq enrichment near two active genes (*Rps19* and *Rps13*) in embryos and somatic tissues (data from refs 9, 17). ATAC-seq peaks near TSSs and TESs are shaded. **c**, Enrichment of repeats in ATAC-seq promoter and distal peaks compared to that in random peaks for early embryos and somatic tissues. The enrichment was calculated as a log₂ ratio for the numbers of observed peaks that overlap with repeats divided by the numbers for random peaks. **d**, Heat maps show the enrichment (calculated similarly as in c) of repeat subfamily in ATAC-seq peaks in early embryos and mESCs.

Accessible chromatin at TESs and repeats

Next, we investigated whether the accessible chromatin landscape in preimplantation embryos is different from those at late developmental stages or in somatic cells. Surprisingly, we observed unusually strong ATAC-seq signals just downstream of transcription end sites (TESs) in 2-cell embryos (Fig. 3a). Such enrichment decreases at late stages and is much weaker in somatic cells^{9,17} (Fig. 3b). The open chromatin at TESs in early embryos preferentially occurs at active genes (Extended Data Fig. 5a, b). The TES open chromatin may reflect the binding of factors engaged in transcription termination¹⁸. Alternatively, these sites may function as enhancers that promote high levels of transcription for housekeeping genes at the onset of ZGA¹⁹. These data revealed that open chromatin in early embryos is found both at promoters and near TESs of active genes.

In addition, we found distal ATAC-seq peaks in early embryos are also distinct from *cis*-regulatory elements in somatic tissues²⁰ (Extended Data Fig. 6a). Notably, certain classes of repeats are highly transcribed in preimplantation embryos²¹ and repetitive elements can function as promoters and enhancers^{22,23}. We confirmed that the 2-cell stage embryos contain the highest fraction of transcripts from repeats among all stages (Extended Data Fig. 6b). In notable contrast to all of the somatic tissues that we examined, both promoter and distal ATAC-seq peaks showed preferential enrichment for repeats at the 2-cell stage (Fig. 3c, d), with several subfamilies of short interspersed nuclear elements (SINEs) (B1, B2, and B4) and endogenous retroviral elements (ERVL) most enriched. Indeed, the repeat-associated peaks at the 2-cell stage probably contributed to the increased total number of ATAC-seq peaks (Extended Data Fig. 6c). In support of a functional role for these transposable elements²¹, the transcription start sites (TSSs) of 2-cell specific genes are more likely to contain repeats particularly for long terminal repeats (LTRs) and SINEs (Extended Data Fig. 6d–f). As validation, repetitive elements are also enriched in H3K27ac peaks in the 2-cell embryos (Extended Data Fig. 6g). Notably, none of the transposable elements that we examined (B1, B2, MERVL, and B4) were

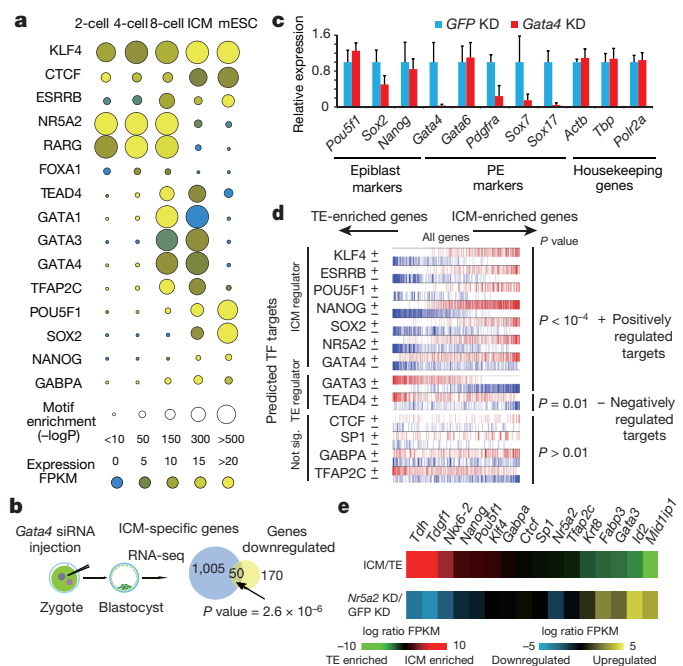


Figure 4 | Identification of candidate regulators for early development.

a, TF motifs identified from distal ATAC-seq peaks. Only TFs expressed at least at one stage (FPKM ≥ 5) and motif enrichment P value $< 1 \times 10^{-10}$ at least at one stage were included. **b**, Schematic of the *Gata4* knockdown experiments. A Venn diagram shows the overlap between *Gata4*-knockdown downregulated genes and ICM-specific genes (compared to mESCs), with the P value (hypergeometric distribution) indicated. **c**, Bar charts showing the relative gene expression with the GFP knockdown (KD) embryos normalized to 1. Error bars denote the standard errors of single-embryo RNA-seq FPKM values ($n = 3$ for GFP KD; $n = 4$ for *Gata4* KD). **d**, Key regulators for ICMs and TEs identified by MARINA. In each row, all genes were sorted (from left to right) by their differential expressions in ICM versus TE cells. Predicted positively and negatively regulated TF targets are marked as red or blue bars on the basis of their co-expression patterns with TFs across individual cells in blastocysts. The P values (FDR-corrected) represent the statistical significance of enrichment estimated by permuting the ICM and TE samples. **e**, Heat maps show ICM to TE expression ratio (top) and *Nr5a2* to GFP knockdown expression ratio in the 8-cell embryos (bottom).

enriched near TESs (Extended Data Fig. 6h), suggesting that the open chromatin near TESs is unlikely to be caused by repeats. In sum, these data suggest that the accessible chromatin landscape in early embryos is extensively shaped by transposable elements.

Regulatory network in early development

As enhancers are known to be hotspots of transcription factor (TF) binding²⁴, we asked whether the distal ATAC-seq peaks harbour motifs for TFs regulating preimplantation development. We confirmed that distal ATAC-seq peaks are highly stage-specific (Extended Data Fig. 7a). The GREAT analysis, which annotates non-coding regions by analysing their nearby genes²⁵, showed that early stage-specific distal peaks are frequently located near genes functioning in chromatin regulation (Extended Data Fig. 7b). Consistently, chromatin regulator genes are generally upregulated at early stages (2–8 cell stages) (Extended Data Fig. 7c). In contrast, distal peaks in ICMs and mESCs are preferentially associated with genes involved in development (Extended Data Fig. 7b). Next, using motif analysis software HOMER²⁶ we found the binding motifs for a set of transcription factors enriched in distal peaks in a highly stage-specific manner (Fig. 4a). Notably, the majority of these transcription factors, including CTCF, NR5A2, TEAD4, GATA4, POU5F1, SOX2 and NANOG, are essential for early development. Importantly, the timing of the appearance of the motifs

coincides with the expression of their corresponding TFs. For example, both the binding motifs and the expression of NR5A2 and RARG are strongly enriched at the 2–8-cell stages (Fig. 4a). A similar observation was made for FOXA1 (4–8-cell stages), POU5F1, SOX2 and NANOG (known regulators of ICMs and mESCs). Interestingly, the GATA factors are strongly enriched in distal peaks in ICMs but not in mESCs (Fig. 4a) (the motifs for individual GATA family members are highly similar, data not shown). Whereas mESCs are typically derived from ICMs, it remains unknown to what extent their chromatin states truly resemble each other. We found that ICMs and mESCs showed distinct landscapes for distal ATAC-seq peaks (Extended Data Fig. 7b) and significant differences in gene expression (Extended Data Fig. 7d, e). As ICMs can give rise to primitive endoderm (PE) and epiblast⁵ but mESCs most resemble preimplantation epiblast²⁷, we speculate that genes expressed in ICMs, but not in mESCs, may preferentially enrich for PE-fate genes. As GATA4, an essential regulator for embryonic development, is a marker of PE²⁸ and is not expressed in mESCs (Fig. 4a), we tested whether the downregulation of GATA4 may promote the conversion of the transcriptome of ICMs towards mESCs. We injected *Gata4* siRNA in zygotes and analysed the effect on blastocysts using RNA-seq (Fig. 4b). Indeed, genes downregulated by *Gata4* knockdown are strongly enriched for ICM-specific genes (Fig. 4b) (overlap gene $n = 50$, P value = 2.6×10^{-6}), but not for mESC-specific genes (overlap gene $n = 24$, P value = 0.68). The expression of PE marker genes *Sox7*, *Sox17* and *Pdgfra*, but not the epiblast markers *Nanog* and *Pou5f1*, was substantially reduced (Fig. 4c). The expression of *Gata6*, which functions upstream of *Gata4* (ref. 29), remained unchanged. Interestingly, *Sox2* is also downregulated upon *Gata4* knockdown (Fig. 4c). As SOX2 is known to promote PE lineage³⁰, we speculate that GATA4 may also in turn regulate the expression of *Sox2*. These data suggest that GATA4 is a regulator of ICM circuitry and its downregulation may have a role in resetting the transcription programs of ICMs to mESCs.

An even earlier lineage specification event before PE and epiblast differentiation is the segregation of ICM and trophectoderm (TE)⁵. To determine which factors regulate ICM and TE segregation, we performed RNA-seq in isolated ICMs and TEs. Interestingly, the known master regulators of ICM (such as *Pou5f1* and *Nanog*) and TE (such as *Cdx2* and *Eomes*) are not among the most differentially expressed genes (Extended Data Fig. 8a). This is possibly because these key regulators are expressed before lineage segregation and their transcripts are only asymmetrically distributed between TEs and ICMs³¹. In contrast, genes exclusively expressed in ICMs but not in TEs are often expressed at later stages (Extended Data Fig. 8a), indicating that they are probably downstream effectors. To identify possible upstream transcription factors, we employed MARINA (MAster Regulator INference algorithm)³², which identifies candidate key TFs by investigating whether the expression of TF target genes (instead of TFs themselves) are enriched in cell-state-specific transcription programs. To identify the possible target genes for TFs identified in our motif analyses, we predicted promoter–enhancer (distal peak) pairs with correlated activities (ATAC-seq enrichment) across development stages as previously described³³ (Extended Data Fig. 8b). The promoter–enhancer pairs were validated by high levels of correlation between corresponding gene expression and ATAC-seq signal on distal peaks (Extended Data Fig. 8c, d). We then searched for the TF motifs in each distal peak and linked the connected promoter and gene to the TFs as their targets (Extended Data Fig. 8b). The TF targets were further separated into positively and negatively regulated genes based on their co-expression patterns with the TFs, using single-cell RNA-seq data in blastocysts¹². By doing so, MARINA correctly identified known ICM markers POU5F1, NANOG, SOX2, ESRRB and KLF4, as well as the TE regulators GATA3 and TEAD4 (Fig. 4d). As controls, factors such as CTCF, GABPA and TFAP2C were not identified as either ICM or TE regulators. We focussed on NR5A2, which is known to promote the expression of *Pou5f1* and *Nanog* and the knockout of *Nr5a2* leads to early embryo death at E6.5–7.5

(refs 34, 35). Interestingly, *Nr5a2* is highly expressed at the 2–8 cell stages (Fig. 4a and Extended Data Fig. 9a), raising the possibility that it may be an early regulator of ICM/TE programs. In support of this, analysis of individual single-cell RNA-seq data¹² showed that in the 8-cell embryos, the expression of *Nr5a2* is positively correlated with the ICM marker genes including *Pou5f*, *Sall4* and *Tdgf1*, and is negatively correlated with the TE marker genes such as *Cdx2* (Extended Data Fig. 9b). This pattern was even more evident in blastocysts. We then knocked down *Nr5a2* in zygotes and collected embryos at the 8-cell stage for RNA-seq analysis (Extended Data Fig. 9c). Indeed, ICM marker genes including *Nanog*, *Pou5f*, *Tdgf1* and *Tdh* were downregulated (Fig. 4e). By contrast, TE marker genes such as *Id2*, *Gata3*, *Fabp3* and *Krt8* were upregulated. TFs that are expressed in both ICM and TE such as *Gabpa*, *Ctcf* and *Sp1* were not affected. Therefore, these data indicate that NR5A2 regulates the expression of key TFs for ICM and TE as early as the 8-cell stage. Notably, the 8-cell-specific genes are also preferentially downregulated upon depletion of *Nr5a2* (Extended Data Fig. 9c). Taken together, by integrating the information about *cis*-regulatory elements and the transcriptome, we uncovered a regulatory network for preimplantation development that is orchestrated by a set of transcription factors and their targets.

A unique chromatin state in minor ZGA

Interestingly, whereas *cis*-regulatory elements are crucial for gene regulation, previous studies suggested that enhancers may be dispensable for transcription of reporters in zygotes³⁶. Instead, enhancers only become essential during the course of ZGA when chromatin is proposed to progressively adopt a repressive state³⁷. However, these experiments were largely based on exogenous reporters. To investigate the chromatin state before major ZGA *in vivo*, we examined early 2-cell embryos in which minor ZGA is most evident but preceding major ZGA³⁸. Our RNA-seq analysis found only a few genes ($n=98$, fragments per kilobase of transcript per million mapped reads (FPKM) > 5) actively transcribed at this stage and a large fraction (48%) of them (such as *Zscan4*) reside in clusters in the genome (Extended Data Fig. 10a–c). Consistent with limited gene activities in early 2-cell embryos, we obtained generally weak and noisy ATAC-seq enrichment, and the detected peaks are reduced for both number (Extended Data Fig. 10d) and genome coverage (Extended Data Fig. 10e) compared to those at the 2-cell stage. These peaks are enriched for repetitive elements with classes similar to those found in 2-cell embryos (Extended Data Fig. 10f). Interestingly, when searching for regions (100 kb bin) with strongest ATAC-seq enrichment in the genome, we discovered that a large fraction of them (52% of the top 100 regions) contain or are in the proximity (100 kb) of the full-length MERVLs. These full-length MERVLs have relatively intact open reading frames and are usually highly transcribed (data not shown). Notably, MERVLs are the most highly transcribed repeats in the 2-cell embryos where they may initiate over 300 genes^{21,39}. The expression of MERVLs was reported to be restricted in a very short time window at the early 2-cell stage⁴⁰, although our RNA-seq data indicate that their transcripts may be accumulated and retained at later stages (Extended Data Fig. 10g). Surprisingly, the open chromatin near MERVL is large in size (up to 117 kb; median length, 40 kb) and is specifically present at the 3' downstream of active MERVLs (Fig. 5a, b, Extended Data Fig. 10h). For brevity, we termed these domains as 'minor-ZGA-associated accessible chromatin domains' (MAC domains). Interestingly, promiscuous transcription lacking 3' processing or splicing is reported to occur specifically in minor ZGA⁴¹. We found that the MAC domains coincide with strong unterminated transcripts of MERVLs that invade the 3' downstream regions (Fig. 5a–c). The MAC domains near MERVLs were absent in either 2-cell embryos or the early 2-cell embryo treated with alpha-amanitin, which blocks transcription (Fig. 5a, b). Finally, we found that non-repeat early 2-cell genes are also preferentially associated with broad ATAC-seq domains, although at weaker levels than

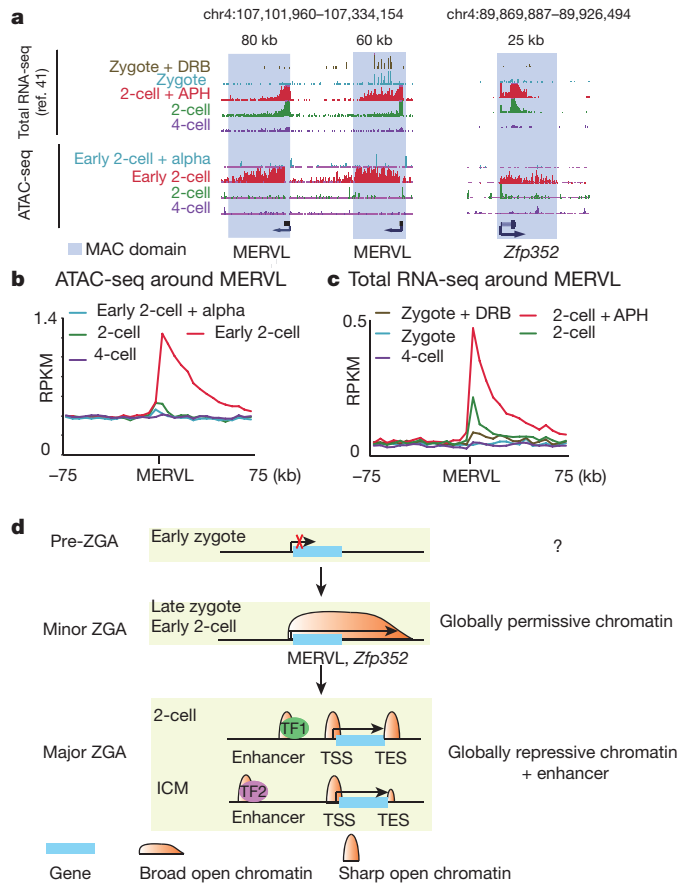


Figure 5 | Transcription regulation and chromatin state in minor ZGA. **a**, The UCSC genome browser views show the promiscuous transcription⁴¹ and broad accessible chromatin domains downstream of MERVL or *Zfp352*. Zygote + DRB, zygote treated with transcription inhibitor DRB; APH, aphidicolin; alpha, alpha-amanitin. **b**, The average ATAC-seq enrichment around MERVLs. The MERVL gene itself is not shown owing to its highly repetitive nature and the resulting low mappability (the upstream and downstream regions of MERVL are mappable). **c**, The average promiscuous transcription levels around MERVLs measured by total RNA-seq⁴¹. **d**, A model shows the distinct transcription and chromatin states in pre-ZGA, minor ZGA and major ZGA.

those near MERVLs (Extended Data Fig. 10i). These include *Zfp352*, which is the most highly expressed non-repeat gene in our data (Fig. 5a and Extended Data Fig. 10j). Taken together, our study showed unique chromatin landscape in minor ZGA that features broad domains of open chromatin covering promiscuous transcription.

Discussion

Here we provide a genome-wide survey of accessible chromatin in the mouse preimplantation embryos using ATAC-seq. A fundamental question in preimplantation development is to what extent gene expression is linked to epigenome reprogramming. Our data indicate that gene activation and establishment of open chromatin could occur, at least in part, through different pathways from those for epigenetic modification reprogramming. One possible reason for the differences is that epigenetic modifications can be inherited from oocyte and sperm. The open chromatin, in contrast, could be newly established when global transcription takes place. Another surprising finding is the unique chromatin state in the minor ZGA featured by broad domains of accessible chromatin over extended promiscuous transcripts (Fig. 5d). Such broad accessible chromatin is in notable contrast to open chromatin at short regulatory elements and supports a globally permissive chromatin state in minor ZGA⁴¹. The relaxed chromatin environment may contribute to the pervasive transcription from repeats and

the ultimate integration of these elements in the host program in evolution³⁹. On the other hand, such promiscuous transcription and globally permissive chromatin may be detrimental for major ZGA. The establishment of repressive chromatin together with stage-specific regulatory elements such as promoters and enhancers probably ensures accurate control of the zygotic transcription program. Intriguingly, transposable elements and the TESs of active genes also show strong ATAC-seq enrichment after major ZGA, particularly at early stages. These findings indicate additional gene regulatory modules in preimplantation embryos. We postulate that the TES accessible chromatin may be associated with regulators actively engaged at TESs to prevent promiscuous transcription, as observed in minor ZGA¹⁸. Alternatively, it may promote efficient and potent transcription by forming loops with the promoters¹⁹ for housekeeping genes. Taken together, our data not only unveiled the chromatin landscapes in minor and major ZGAs, but also allowed genome-wide identification of regulatory circuitry in early development. Further investigations with additional complementary approaches are warranted to fully dissect epigenomic reprogramming during preimplantation embryogenesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 October 2015; accepted 27 May 2016.

Published online 15 June 2016.

- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
- Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).
- Burton, A. & Torres-Padilla, M. E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 723–735 (2014).
- Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* **10**, 467–477 (2009).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
- Cusanovich, D. A. *et al.* Epigenetics. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
- Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
- Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
- Gu, W. *et al.* Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **17**, 41 (2016).
- Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
- Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
- De Braekeleer, E. *et al.* ETV6 fusion genes in hematological malignancies: a review. *Leuk. Res.* **36**, 945–961 (2012).
- Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**, 21–32 (2001).
- Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
- Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769–774 (2015).
- Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–78 (2008).
- Randise-Hinchliff, C. E. & Brickner, J. H. A new direction for gene looping. *Dev. Cell* **23**, 919–921 (2012).
- Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
- Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
- Robbez-Masson, L. & Rowe, H. M. Retrotransposons shape species-specific embryonic stem cell gene expression. *Retrovirology* **12**, 45 (2015).
- Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* **45**, 836–841 (2013).
- Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
- McLean, C. Y. *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 516–528 (2014).
- Soudais, C. *et al.* Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation *in vitro*. *Development* **121**, 3877–3888 (1995).
- Morrissey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12**, 3579–3590 (1998).
- Wicklow, E. *et al.* HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS Genet.* **10**, e1004618 (2014).
- Jedrusik, A., Cox, A., Wicher, K. B., Glover, D. M. & Zernicka-Goetz, M. Maternal-zygotic knockout reveals a critical role of Cdx2 in the morula to blastocyst transition. *Dev. Biol.* **398**, 147–152 (2015).
- Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
- Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
- Labelle-Dumais, C., Jacob-Wagner, M., Paré, J. F., Bélanger, L. & Dufort, D. Nuclear receptor NR5A2 is required for proper primitive streak morphogenesis. *Dev. Dyn.* **235**, 3359–3369 (2006).
- Gu, P. *et al.* Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell Biol.* **25**, 3492–3505 (2005).
- Schultz, R. M. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum. Reprod. Update* **8**, 323–331 (2002).
- Aoki, F., Worrall, D. M. & Schultz, R. M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev. Biol.* **181**, 296–307 (1997).
- Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131 (2004).
- Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Ishichi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
- Abe, K. *et al.* The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing. *EMBO J.* **34**, 1523–1537 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We appreciate B. Ren, D. Leung, H. Yang and the members of the Xie laboratory for comments during preparation of the manuscript. This work is supported by the funding provided by the National Basic Research Program of China (973 program) 2015CB856201 (W.Xie), the National Natural Science Foundation of China 31422031 (W.Xie), 31171381 (J.N.), 81472855 (X.Y.), the National Basic Research Program of China 2012CB966701 (J.N.), the Beijing Natural Science Foundation grant 5152014 (J.N.), Tsinghua University Initiative Scientific Research Program (20131089278, 2014z21046) (X.Y.), the funding from the THU-PKU Center for Life Sciences (W.Xie, X.Y.), and the Youth Thousand Scholar Program of China (W.Xie, X.Y.).

Author Contributions J.W., B.H. and W.Xie conceived and designed the experiments. J.W. conducted the ATAC-seq experiments. J.W., B.L., W.Xia and Q.W. developed CARM. B.H. performed the mouse embryo experiments with help from Y.X., J.M., W.L. and J.Z. H.C. prepared the RNA-seq libraries. Q.Y. and W.Z. helped with various experiments. J.W., H.C., H.Z., Y.L., X.Y. and W.Xie performed the bioinformatics analysis of the data. B.Z. conducted the ChIP-seq experiment. X.P., F.X., G.T. advised the development or application of ATAC-seq and CARM. Y.L. and Q.W. performed NGS sequencing. Z.C. and J.N. supervised the mouse work. J.W., H.C., B.H. and W.Xie wrote the manuscript.

Author Information All data have been deposited to GEO with the accession number GSE66390. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.Xie (xiewei121@tsinghua.edu.cn).

Reviewer Information Nature thanks T. Magnuson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during outcome assessment.

Embryo collection. All analysed embryos were collected from 5- to 6-week-old C57BL/6N females mated with DBA/2N males (Vital River). For reciprocal experiments, C57BL/6N males and DBA/2N females were used. To induce ovulation, females were administered 5 IU of hCG intraperitoneally, 44–48 h after injection of 5 IU of PMSG (San-Sheng pharmaceutical Co. Ltd). Each set of embryos at a particular stage was flushed from the reproductive tract at defined time periods after hCG administration: 30 h (early 2-cell), 39–43 h (2-cell), 54–56 h (4-cell), 68–70 h (8-cell) and 92–94 h (blastocysts) in HEPES-buffered CZB medium. To inhibit transcription in the minor ZGA, early zygotes (PN3) were cultured in CZB supplemented with alpha-amanitin for about 14 h. Embryos were selected by cell number or morphology, with their zona pellucida gently removed by treatment of 10 IU ml⁻¹ pronase (Sigma P8811) for several minutes. Blastocysts were incubated in a 1:3 dilution of anti-mouse rabbit serum in DMEM medium for 20 min, washed in PBS and further incubated for 20 min in a 1:5 dilution of rat serum in DMEM for the complement reaction. The ICM was subsequently cleaned from lysed trophectoderm with a narrow glass pipette. The resulting cells were manually picked and treated with the lysis buffer for ATAC-seq or Smart-seq2.

All animal maintenance and experimental procedures were carried out according to guidelines of Institutional Animal Care and Use Committee (IACUC) of Tsinghua University, Beijing, China.

Cell culture. mESCs (R1) were cultured on gelatin in DMEM containing 15% FBS, leukaemia inhibiting factor, penicillin/streptomycin, L-glutamine, β-mercaptoethanol, and non-essential amino acids.

ATAC-seq library preparation and sequencing. The ATAC-seq libraries of mESCs and early mouse embryos were prepared as previously described with minor modifications⁶. Briefly, samples were lysed in lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂ and NP-40) for 10 min on ice to prepare the nuclei. The optimized concentration of NP-40 is 0.15% for mESCs, the 2-cell, 4-cell, 8-cell embryos and 0.5% for early 2-cell and ICMs. Immediately after lysis, nuclei were spun at 500g for 5 min to remove the supernatant. Nuclei were then incubated with the Tn5 transposome and tagmentation buffer at 37 °C for 30 min (Vazyme Biotech). After the tagmentation, the stop buffer was added directly into the reaction to end the tagmentation. PCR was performed to amplify the library for 15 cycles using the following PCR conditions: 72 °C for 3 min; 98 °C for 30 s; and thermocycling at 98 °C for 15 s, 60 °C for 30 s and 72 °C for 3 min; following by 72 °C 5 min. After the PCR reaction, libraries were purified with the 1.2× AMPure (Beckman) beads before proceeding for mitochondrial DNA depletion.

CRISPR/Cas9-assisted removal of mitochondrial DNA (CARM). The NGG protospacer adjacent motif (PAM) sequences were first identified in the genome of mitochondria, and the upstream 20 bp of PAM sequences was used as sgRNA candidates. A total of 114 sgRNAs were then selected and synthesized to cover approximately every 140 bp over the mitochondrial genome. Paired oligonucleotides were annealed, pooled, and were ligated to the pUC57kan-T7-gRNA vector, which were further transformed and amplified. *In vitro* transcription was performed to produce sgRNAs (MEGashortscript™ Kit, Thermo Fisher Scientific). Each ATAC-seq library was incubated with 330–500 ng sgRNA and 1 μg Cas9 protein (PNA Bio CP01-50) for 2 h at 37 °C. After incubation, the reaction was treated by RNaseA before being terminated by adding the stop buffer (30% glycerol, 1.2% SDS, 250 mM EDTA, pH 8.0). The ATAC-seq library was further purified by 1.2× AMPure beads and was subjected to sequencing on HiSeq1500 or 2500 (Illumina) according to the manufacturer's instruction.

RNA-seq library preparation and sequencing. The RNA-seq libraries were generated from early mouse embryos and ES cells using Smart-seq2 (ref. 42) with minor modification. Cells were lysed in hypotonic lysis buffer (Amresco, M334), and the polyadenylated mRNAs were captured by the PolyT primers. After ~3–10 min lysis at 72 °C, the Smart-seq2 reverse transcription reactions were performed. After pre-amplification and AMPure XP beads purification, cDNAs were sheared by Covaris and were subject to Illumina TruSeq library preparation. All libraries were sequenced on Illumina HiSeq1500 or 2500 according to the manufacturer's instruction.

ChIP-seq library preparation and sequencing. To investigate histone modifications in early embryos, we developed a low-input ChIP-seq method (manuscript in submission). Briefly, each sample is lysed and subjected to MNase digestion at 37 °C. The reaction is terminated by adding stop buffer (110 mM Tris-HCl (pH 8.0), 55 mM EDTA) and cold 2× RIPA buffer. After spinning at max speed, the supernatant is transferred to a new tube. Each chromatin sample is supplemented with RIPA buffer and is incubated with antibodies for H3K27ac (Active Motif 39133)

or H3K27me3 (Diagenode pAb-069-050) (overnight at 4 °C). The next day, the sample is incubated with protein A dynabeads (Life Technologies) for 2 h with rotation at 4 °C. The beads are washed with RIPA buffer and LiCl buffer. After washing, tubes are spun briefly and the supernatant is removed. Beads are resuspended with ddH₂O and Ex-Taq buffer (TaKaRa). Proteinase K (Roche) is then added to elute DNA from beads. The supernatant is then transferred to a new tube and the proteinase K is heat inactivated. The end of DNA is repaired by rSAP (NEB) followed by heat inactivation. The resulting sample is subjected to library preparation as previously described⁴³.

Gene knockdown in mouse embryos. Fertilized C57BL/6J zygotes for siRNA microinjection were collected from the oviducts of female mice at 18 h after injection with hCG. All siRNA oligonucleotides were synthesized with 5'-CY3 and 2'-O-Me modification. The siRNA sequences were as follows: Gata4-Mus-1748: 5'-GUCCCAGACAUCAGUACUTT-3'; Gata4-Mus-1244: 5'-GGCAGAGAGUGUGUCAUUTT-3'; Gata4-Mus-2811: 5'-GGUUGUGUCUACAGCACAATT-3'; Nr5a2-Mus: 5'-GCAAGTGTCTAATTAAAA-3'. An injection pipette containing 20 mM siRNA solution was inserted into the cytoplasm of the zygotes. About 40 zygotes were microinjected and further cultured up to the 8-cell (Nr5a2) or blastocyst stage (Gata4) in CZB medium containing glutamine at 37 °C under 5% CO₂ in air.

ATAC-seq data processing. All ATAC-seq reads were first aligned to the genomes of the C57BL and DBA strains separately using Bowtie2 (version 2.2.2). SNP tables for DBA/2J and C57BL/6N were downloaded from the Sanger Institute Mouse Genome Project⁴⁴. The DBA/2J and C57BL/6N genomes were generated by substituting corresponding bases from the mm9 genome. As we used the DBA/2N strain instead of the DBA/2J strain, we verified the identity of this strain by sequencing its genome. The genomes of DBA/2N and DBA/2J are highly similar as 99.4% of SNPs identified in the DBA/2N strain (compared to the reference genome) are the same as those found in the DBA/2J strain identified by the Sanger Institute (data not shown). Therefore, we used the DBA/2J genome (which has a deeper sequencing depth) for all subsequent analyses. The single-end ATAC-seq reads were aligned with the parameters -t -q -N 1 -L 25 and the paired-end ATAC-seq reads were aligned with the parameters: -t -q -N 1 -L 25 -X 2000 no-mixed no-discordant. ChIP-seq reads were aligned to mm9 with the parameters -t -q -N 1 -L 25. All unmapped reads, non-uniquely mapped reads and PCR duplicates were removed. For downstream analysis, we normalized the read counts by computing the numbers of reads per kilobase of bin per million of reads (RPKM). To visualize the ATAC-seq signal in the UCSC genome browser, we extended each read by 250 bp and counted the coverage for each base.

RNA-seq data processing. All RNA-seq data were mapped to the mm9 genome by STAR (version 2.4.0)⁴⁵, which was shown to be highly effective in mapping RNA-seq reads containing SNPs¹². The RNA-seq data for mESCs and ICMs from Tang *et al.*⁴⁶ were downloaded and mapped similarly. The gene expression level was calculated by Cufflinks (version 2.2.1) using the refFlat database from the UCSC genome browser. To quantify the expression level of total repeats in mouse preimplantation embryos, the reads were mapped to RepBase by Bowtie2 (version 2.2.2) and only uniquely mapped reads were kept for further analysis. To obtain reliable reads counts for each family of repeats (MERVL), reads mapped to mm9 were counted as RPKM on the basis of the locations of annotated repeats (RepeatMasker) downloaded from the UCSC genome browser.

Allele assignment of sequencing reads. Allele assignment of sequencing reads was conducted as described previously⁴⁷ and we validated our allelic analyses using the sequenced genomic DNA (cortex) extracted from the C57BL/6N and DBA/2N strains used in this study, which showed that 99.6% and 98.9% reads were correctly assigned, respectively.

The comparison between ATAC-seq replicates. The correlation between different ATAC-seq replicates was calculated as following: each read was extended 250 bp from the mapped end position and the RPKM value was generated on a 100-bp-window base. The ATAC-seq enrichment was then summed within each 5-kb-window for the entire genome and was compared across different samples. Pearson correlation was used for all analyses.

Identification of promoter and distal ATAC-seq peaks. All the ATAC-seq peaks were called by MACS with the parameters nolambda -nomodel. ATAC-seq peaks that are at least 2.5 kb away from annotated promoters (RefSeq, Ensemble and UCSC Known Gene databases combined) were selected as distal ATAC-seq peaks.

Comparison of ATAC-seq peaks and known cis-regulatory elements or H3K27ac ChIP-seq peaks. To compare the ATAC-seq peaks identified in early embryos with the annotated cis-regulatory elements, we calculated the overlap between the ATAC-seq peaks of different stages and ± 0.5 kb around the TSSs of annotated promoters (RefSeq, Ensemble and UCSC Known Gene). Non-promoter (distal) peaks were then compared to distal DHSs in mESCs (ENCODE) or 2-cell H3K27ac ChIP-seq peaks called by MACS. Random peaks were generated by

selecting random regions in the genome with the sizes matching each individual ATAC-seq peak.

Identification of stage-specific genes. A Shannon-entropy-based method was used to identify stage-specific genes, as previously described³³. Due to the possible confounding effects from maternally inherited RNA transcripts, ZGA-only genes were analysed, which were defined as those not expressed in oocytes (FPKM ≤ 0.5) but that are activated (FPKM > 1) after ZGA (in either the 2-cell, 4-cell, 8-cell embryos or ICMs). Genes with entropy score less than 2 were selected as candidates for stage-specific genes. Among these genes, we selected candidates of stage-specific genes for each stage based on the following criteria: the gene is highly expressed at this stage (FPKM ≥ 3), and such high expression cannot be observed in more than two additional stages. These genes were then reported in the final stage specific gene lists.

Identification of ICM and mESC specific genes. ICM- and mESC-specific genes were identified as follows: a minimal of twofold change in expression levels (FPKM) between ICMs and mESCs was required. Genes with low expression (FPKM < 1) in both ICMs and mESCs were removed. Those that fulfil such criteria in both our dataset and that of Tang *et al.*⁴⁶ were selected as the final list of ICM- and mESC-specific genes.

Analysis of the promoter ATAC-seq enrichment. To ensure that the promoters that we analysed reflect the truly active promoters (as many genes have alternative promoters), we only included genes that contain a single promoter, or multiple promoters that are located within 500 bp of each other, in which case a promoter would be randomly chosen. The ATAC-seq signals at promoters (± 2.5 kb) were computed across all promoters in the genome.

Identification of allele-specific accessible chromatin regions. To identify the allele-specific accessible chromatin regions, the mouse genome was divided into consecutive 1-kb-bins and only those bins covered by at least 8 reads of which the parental origins were determined by SNPs were considered. The significance of allele bias for each bin was then assessed via the binomial test and the allele score (AS score) was defined as $-\log_{10}(P \text{ value})$. To estimate the FDR for all reads that contain informative SNPs, we permuted their parental origins by randomly assigning them to two arbitrary alleles. A similar random AS score (R-AS score) for the allele bias was computed for each region using the permuted data set. The permutation was performed five times and the average R-AS scores were used to assess the global FDR under an AS score cutoff. Allele-specific accessible chromatin regions were identified using a cutoff of AS score 3 (absolute value, corresponding to a P value of 0.001), resulting in a less than 2% FDR for all the stages.

Identification of allele-specific genes and ATAC-seq peaks. To identify genes that are expressed in an allele-specific manner, allelic reads mapped to either maternal or paternal allele were identified as described above. Allelic reads mapped to exons for each gene were counted by Htseq-count⁴⁸. Only genes with FPKM ≥ 1 and are covered by at least 30 allelic reads were kept for downstream analysis. The allele-specific genes were identified by at least threefold change between the numbers of maternal and paternal reads and a minimal of AS score of 3 (similarly computed as for the ATAC-seq data). ATAC-seq peaks called by MACS were identified as allele-specific ones if they overlap with 1 kb allele-specific region with at least 50% of the length of the peak. Biallelic ATAC-seq peaks were identified if they overlap with the non-allele-specific 1 kb region covered by SNPs.

DNA methylation analysis at ATAC-seq peaks. The average DNA methylation levels within the defined ATAC-seq peak regions, as well as 7.5 kb of their upstream and downstream regions, were computed. The MethylC-seq data (which measure 5mC + 5hmC), fCAB-seq data (5fC + 5mC), TAB-seq data (5hmC) were downloaded from a previous study¹⁶.

Gene ontology analysis. The DAVID web-tool was used to identify the GO terms⁴⁹.

The comparison between ATAC-seq peaks and repetitive elements. To identify the overlap between repetitive elements and promoter or distal ATAC-seq peaks, the ATAC-seq peaks were compared with the locations of annotated repeats (RepeatMasker) downloaded from the UCSC genome browser. As repeats of different classes vary greatly in numbers, a random set of peaks with identical lengths of ATAC-seq peaks was used for the same analysis as a control. The numbers of observed peaks that overlap with repeats were compared to the numbers of random peaks that overlap with repeats, and a log ratio value was generated as the 'observed to expected' enrichment.

Comparison of stage specificity between promoter and distal ATAC-seq peaks. To compare the stage-specificity of distal ATAC-seq peaks and promoter ATAC-seq peaks across early developmental stages, the Shannon entropy score for normalized RPKM for each peak was calculated across five stages and cell types. Weak ATAC-seq peaks (RPKM < 0.5 at any stages) were removed from this analysis.

Comparison between distal ATAC-seq peaks and somatic tissue enhancers. To compare the distal ATAC-seq peaks identified in early embryos with the annotated

somatic tissue enhancers²⁰, the overlap between the distal ATAC-seq peaks in early mouse embryos and somatic tissue enhancer data sets were calculated.

Identification of stage-specific distal ATAC-seq peaks. The distal ATAC-seq peaks from embryos of all stages and mESCs were combined, with overlapped peaks merged. The average RPKM values were calculated for these distal ATAC-seq peaks which were further normalized by the Z-score normalization. A Shannon-entropy-based method⁵⁰ was used to identify stage-specific distal ATAC-seq peaks. We selected those with entropy less than 2 as candidates for stage specific distal ATAC-seq peaks. The stage-specific distal ATAC-seq peaks were further defined based on the following criteria: the distal ATAC-seq peak is highly active in this stage (normalized RPKM > 1), and its activity (normalized RPKM > 0) cannot be observed at more than two additional stages. The resulting distal ATAC peaks were then reported in the final stage-specific distal ATAC-seq peak list. The functional enrichment for genes that are near stage-specific distal ATAC-seq peaks was analysed using the GREAT tool by default settings²⁵.

Prediction of promoter targets of putative enhancers. To identify the potential targeted genes for each putative enhancer (distal peaks), we computed the averaged ATAC-seq enrichment (normalized RPKM) for all distal ATAC-seq peaks and annotated promoters (TSS ± 2.5 kb). We calculated the correlation across the 5 cell types and stages between the ATAC-seq enrichment at distal ATAC peaks and each promoter within 100 kb. The promoter with a Pearson correlation coefficient of more than 0.8 was selected as the potential target of the enhancer. As a control, for each promoter–enhancer pair, we permuted promoters and putative enhancers pair and generated a random promoter–enhancer pair of equal distance.

Connecting TFs to target genes through distal peak-promoter interactions. To find the sequence motif enriched in distal ATAC-seq peaks, findMotifsGenome.pl from the HOMER program²⁶ was used. AnnotatePeaks.pl was then used to identify specific peaks that contain certain motifs. To connect TFs to genes at a particular developmental stage, we first identified TF motifs that are present in distal ATAC-seq peaks and selected those that are highly enriched (P value $< 1 \times 10^{-10}$). We then assigned these TF motifs to genes by the distal-promoter peak pairs established in this work. If multiple putative enhancers were assigned to the same promoter, the numbers of TF motifs within these enhancers assigned to a gene were then summed. If a gene receives no assignment of motifs for a TF from its linked enhancers, the number of TF motif for this gene is 0.

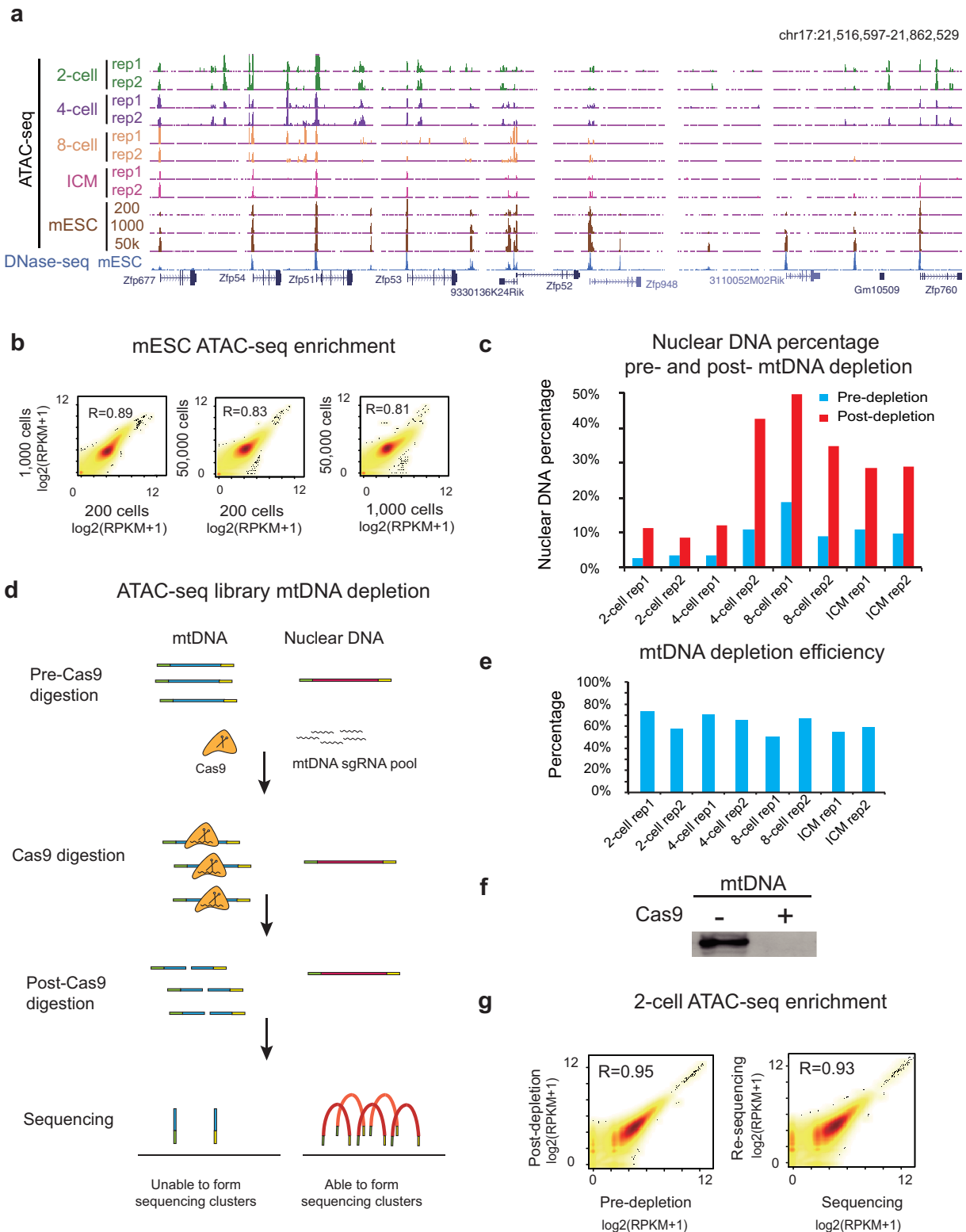
Inference of master regulators between ICM and TE with MARINA. Master regulator analysis was performed using MARINA³². Specifically, we first identified a total of 11 representative ICM and 10 representative TE single cells with high levels of *Pou5f1* transcripts but low *Cdx2* expression, or high *Cdx2* but low *Pou5f1* expression, respectively, using a published single-cell transcriptome data set¹². These single cells were considered as replicates of ICM and TE for identifying ICM- and TE-enriched genes with consideration of their expression variations. Next, for the candidate TFs identified in ICM by HOMER, their target genes were predicted based on TF motif enrichment in promoter-interacting enhancer regions as described above. The positively regulated (or negatively regulated) targets were defined as those genes of which the expression shows positive (or negative) correlation with that of the TF across all individual cells from blastocysts. For each TF, MARINA then assesses enrichment of the predicted targets in the differentially expressed genes between ICM and TE. If the positively regulated and negatively regulated targets are enriched for ICM-specific genes and TE-specific genes, respectively, the corresponding TF is defined as a candidate ICM regulator. Conversely, if the positively regulated and negatively regulated targets are enriched for TE-specific genes and ICM-specific genes, respectively, the TF is defined as a candidate TE regulator. Statistical significance of such enrichment was estimated by comparing the enrichment score with those from the same analysis but after sample permutations for 10,000 times. TFs with FDR-corrected P values ≤ 0.01 were considered significant and thereby inferred as master regulators between ICM and TE.

Selection of downregulated genes for the *Gata4* and *Nr5a2* knockdown embryos. Genes that showed twofold downregulation upon *Gata4* knockdown (average of four single embryos) compared to *GFP* knockdown (average of three single embryos) or upon *Nr5a2* knockdown (average of four single embryos) compared to *GFP* knockdown (average of four single embryos) were selected. Genes with low expression in both *GFP* and knockdown embryos (FPKM < 1) were removed.

Analysis of total RNA-seq data. Total RNA-seq data were downloaded from Abe *et al.*⁴¹. A RPKM file was generated using 100-bp-bin for each data set and was visualized in the UCSC genome browser. To specifically investigate the promiscuous intergenic transcription, exonic transcripts (usually RPKM > 2) from genes need to be excluded, including those inherited from oocytes and those derived from genes activated after the major ZGA. Therefore, we excluded bins with RPKM values above 2 from each total RNA-seq data and the resulting data were used for downstream analysis.

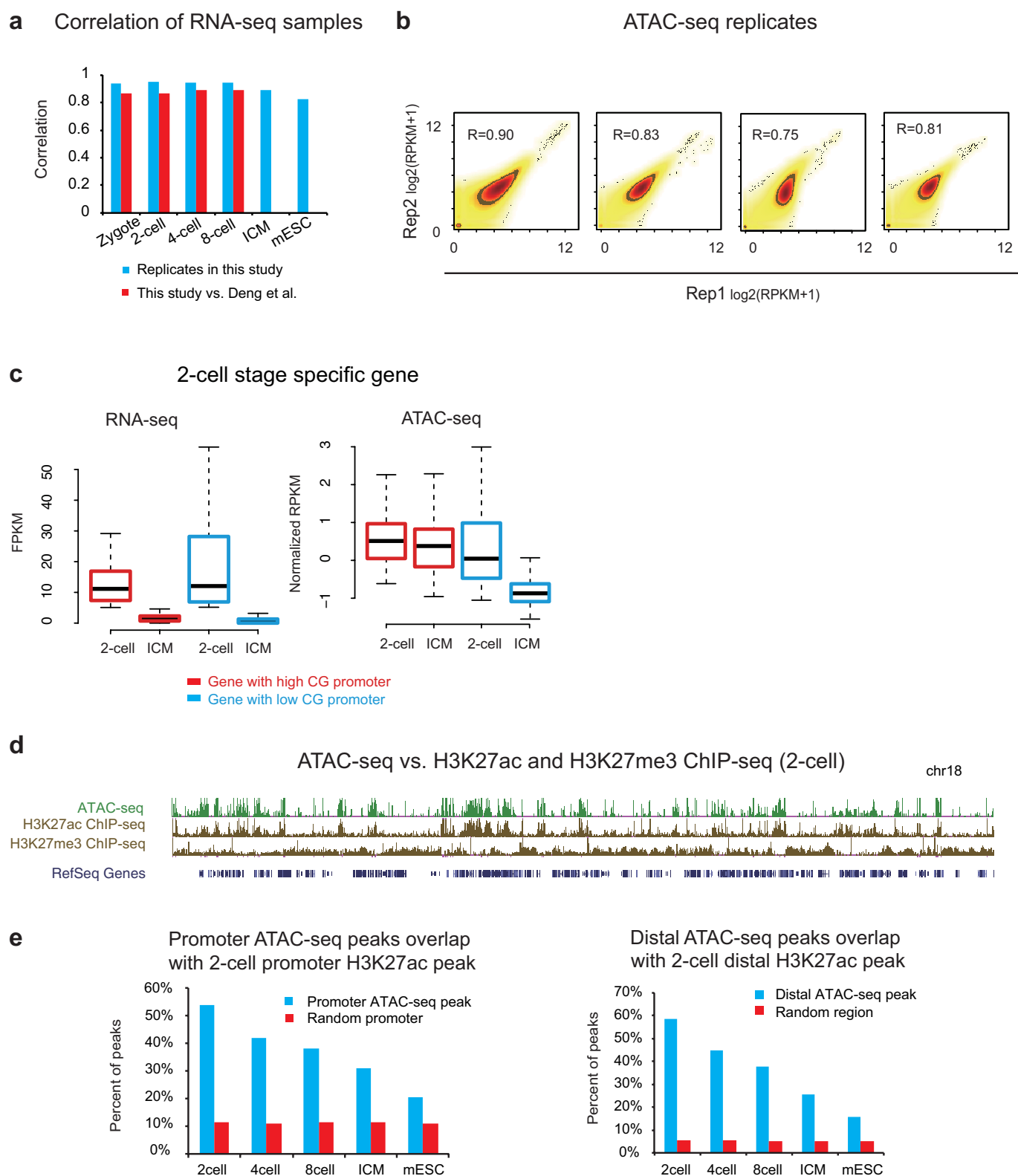
Identification of the broad domains of open chromatin in the early 2-cell embryos. To identify all possible broad domains of open chromatin in the genome, we merged peaks within 50 kb and those with sizes more than 10 kb were identified as broad domains.

42. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
43. Peng, X. *et al.* TELP, a sensitive and versatile library construction method for next-generation sequencing. *Nucleic Acids Res.* **43**, e35 (2015).
44. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
45. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
46. Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* **6**, e21208 (2011).
47. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012).
48. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
49. Dennis, G., Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
50. Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).



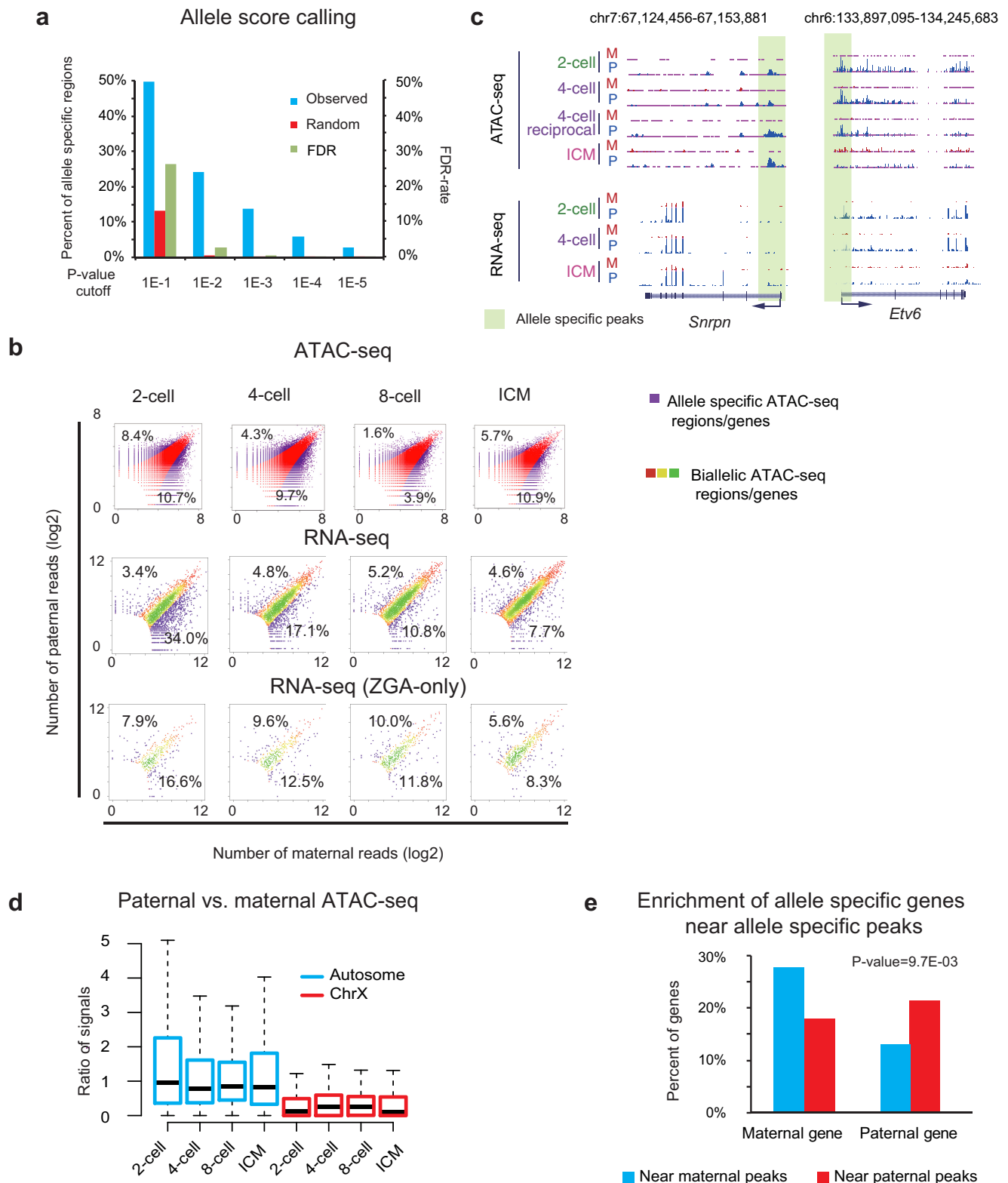
Extended Data Figure 1 | Development of CARM, a method to deplete mitochondrial DNA from ATAC-seq library. **a**, A snapshot of the UCSC browser view shows enrichment of ATAC-seq (with replicates) in early embryos and mESCs, as well as the enrichment of DNase-seq in mESCs. **b**, Scatter plots comparing the ATAC-seq enrichment (RPKM, 5-kb-window for the entire genome) between samples using various numbers of mESCs. The Pearson correlation coefficients are shown. **c**, Bar charts showing the average percentages of monoclonal nuclear DNA reads in ATAC-seq sequencing libraries before and after mitochondrial DNA (mtDNA) depletion for early mouse embryos at each developmental stage. **d**, Schematic of mtDNA depletion for the ATAC-seq library. mtDNA in ATAC-seq library was digested by Cas9 with sgRNAs targeting

mtDNA. These digested mtDNA cannot form sequencing clusters during sequencing due to the lack of adaptors on both ends. **e**, Bar chart shows the mtDNA depletion efficiency for samples of different stages. The depletion efficiency refers to the percentages of total mtDNA reads that were removed by CARM. **f**, The DNA gel showing the Cas9 digestion efficiency for mtDNA. A total of sixteen 1 kb mtDNA amplicons covering the entire mitochondrial genome were pooled and subjected to Cas9 digestion together with mtDNA sgRNAs. **g**, Scatter plots showing the correlation of ATAC-seq enrichment for the 2-cell embryos between pre-depletion and post-depletion. A similar analysis was performed between sequencing and re-sequencing results for the same library as a control.



Extended Data Figure 2 | Validation of ATAC-seq and RNA-seq data in early mouse embryos. **a**, The Spearman correlation between the replicates of RNA-seq samples and between RNA-seq in this study and Deng *et al.*¹² at stages when available. **b**, Scatter plots comparing the enrichment (5-kb window for the entire genome) between ATAC-seq replicates for mouse preimplantation embryos. The Pearson correlation of the RPKM values is shown. **c**, Box plots comparing the levels of gene expression and promoter ATAC-seq enrichment in the 2-cell embryos and ICM for genes with

different promoter CG densities. Genes specifically expressed at the 2-cell stage were analysed. **d**, A snapshot of the UCSC genome browser shows the global view of ATAC-seq, H3K27ac and H3K27me3 ChIP-seq enrichment in the 2-cell embryos. **e**, The overlaps between ATAC-seq peaks of early embryos at each stage and H3K27ac ChIP-seq peaks in the 2-cell embryos for promoter peaks (left) or distal peaks (right). The overlaps of random promoters or peaks and H3K27ac ChIP-seq peaks in the 2-cell embryos are included as controls.

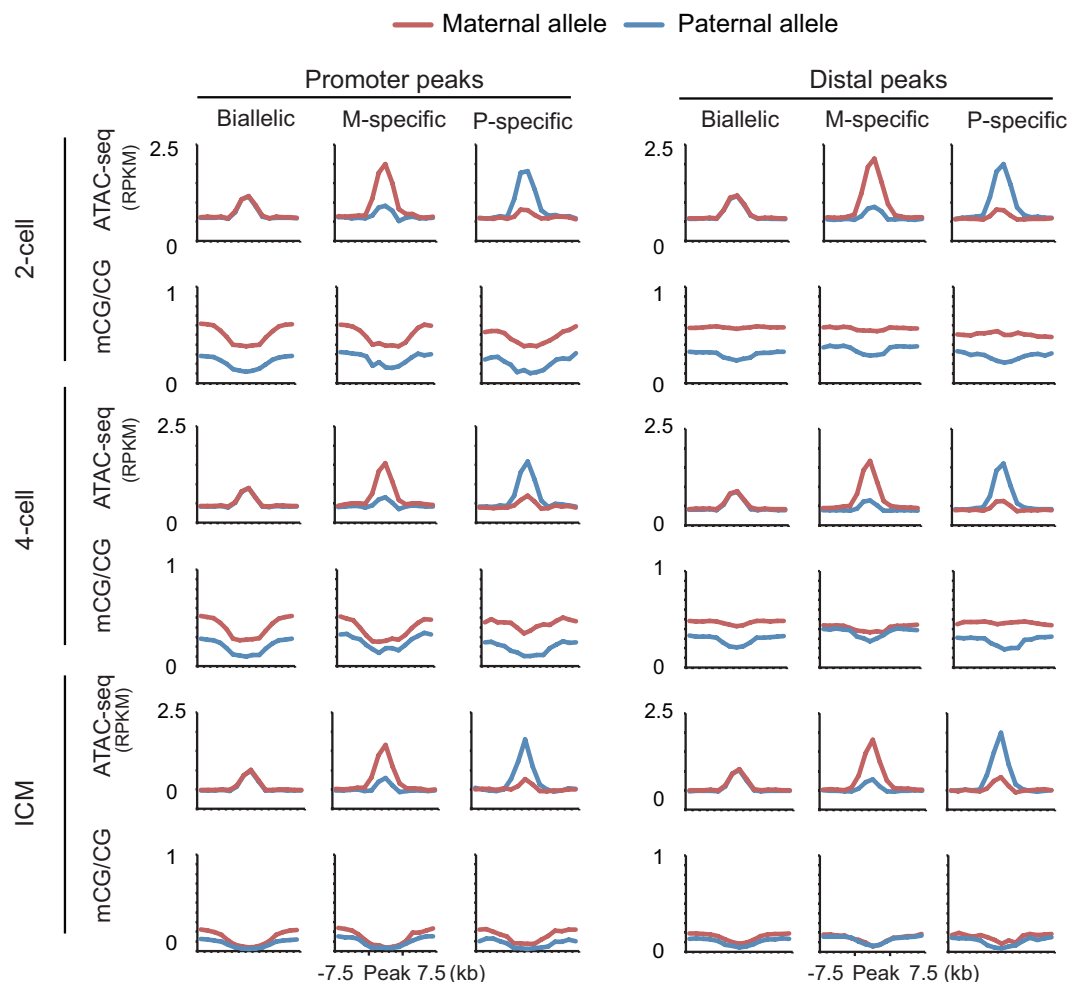


Extended Data Figure 3 | Allelic accessible chromatin in early mouse embryos. **a**, The percentages of regions (covered by at least 8 reads with assigned parental origins) defined as allele specific regions (blue) at various P -value cutoffs are plotted. As a control, these reads were randomly assigned to either allele and a similar percentage was calculated (red, average of 5 random permutations). A P value of 0.001 was selected as the cutoff to give the FDR (random to observed; green) less than 2% for each stage. **b**, Scatter plots showing the numbers of ATAC-seq reads (top) assigned to each allele in SNP-containing regions and RNA-seq reads assigned to each allele for SNP-containing genes (middle and bottom). For RNA-seq analysis, either all genes (middle) or ZGA-only genes (bottom) were used. The purple denotes allele-specific ATAC-seq regions or allele-specific genes. The red, yellow and

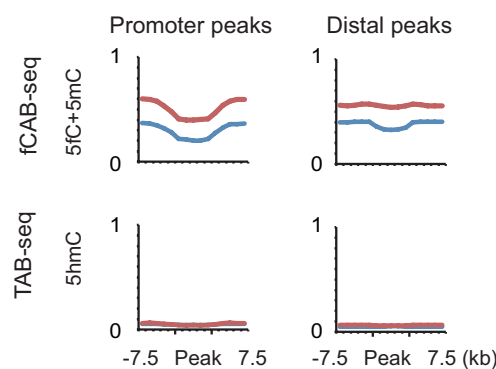
green denote biallelic ATAC-seq regions or biallelic genes, with green to red showing high to low point densities. The numbers show the percentages of maternal- and paternal-specific ATAC-seq regions or genes. **c**, Snapshots show the allelic ATAC-seq signals and RNA-seq signals near the imprinted locus *Snrpn* and near the gene *Etv6*. Note, not all regions are covered by SNPs. **d**, Box plots showing the ratio of paternal versus maternal read numbers for all regions in the genome (1-kb window) covered by at least 8 reads with assigned parental origins. The analysis was performed for either autosomes (blue) or the X chromosome (red). **e**, Bar charts showing the percentages of allele specific genes (combining all stages) that contain allele specific ATAC-seq peaks at the same stages in nearby regions (within 40 kb). A P value based on the hypergeometric distribution is shown.

a

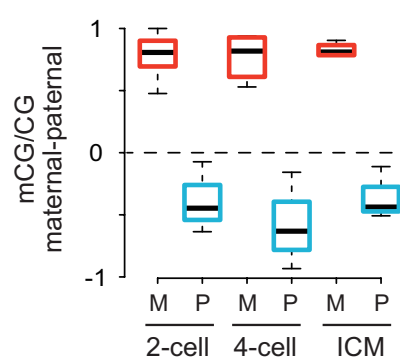
ATAC-seq enrichment and DNA methylation at ATAC-seq peaks

**b**

5fC and 5hmC levels at ATAC-seq peaks (2-cell)

**c**

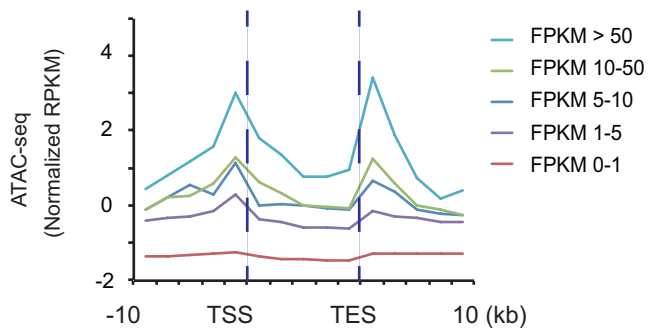
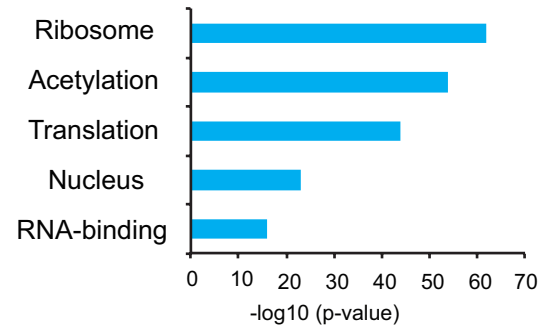
Imprinted regions



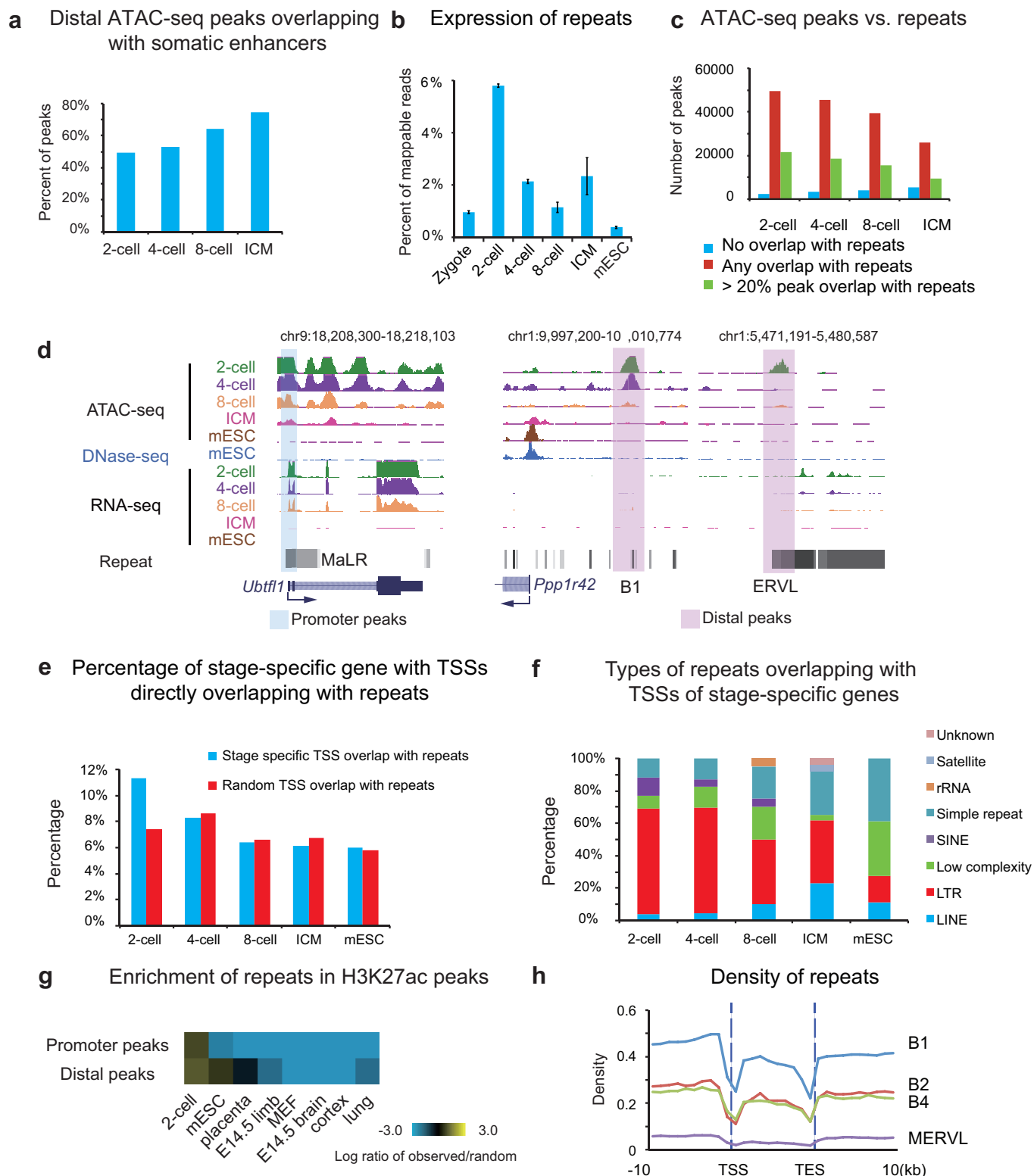
— Maternal allele — Paternal allele

Extended Data Figure 4 | Allelic ATAC-seq enrichment and DNA methylation levels at ATAC-seq peaks. **a**, The allelic average ATAC-seq enrichment and DNA methylation levels around biallelic, maternal- and paternal-specific ATAC-seq peaks are shown for both promoter peaks (left) and distal peaks (right). M-specific, maternal specific; P-specific, paternal-specific. **b**, The allelic average fCAB-seq enrichment, which measures the sum of 5fC and 5mC, and TAB-seq enrichment, which measures 5hmC, are shown for both promoter peaks (left) and distal

peaks (right) in the 2-cell embryos. The fCAB-seq and TAB-seq data were obtained from a previous study¹⁶. Notably, MethylC-seq measures the sum of 5mC and 5hmC. As the levels of 5hmC on both alleles are much lower compared to 5mC, the presence of 5hmC should not significantly alter the total 5mC levels. **c**, Box plots showing allelic DNA methylation differences (maternal - paternal) at the known germline imprinted regions⁴⁷. M, maternal; P, paternal.

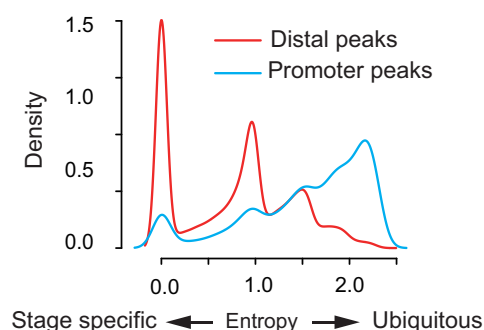
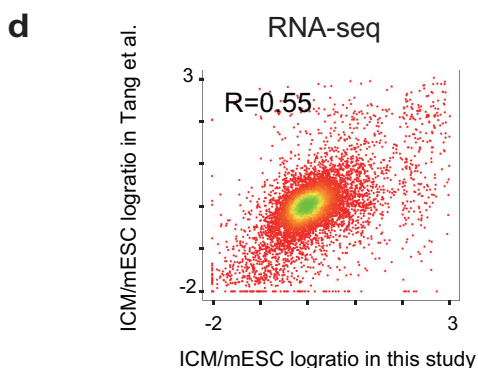
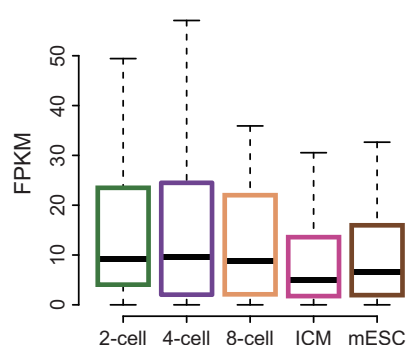
a ATAC-seq enrichment vs. gene expression (2-cell)**b** GO for top 1,000 genes ranked by TES ATAC-seq enrichment (2-cell)

Extended Data Figure 5 | Analyses of ATAC-seq peaks at transcription end sites. **a**, The average levels of ATAC-seq enrichment for genes of different expression levels (ZGA-only) in the 2-cell embryos. **b**, The Gene Ontology analysis results for the top 1,000 genes ranked by TES ATAC-seq enrichment (average enrichment from TES to 2 kb downstream) in the 2-cell embryos.

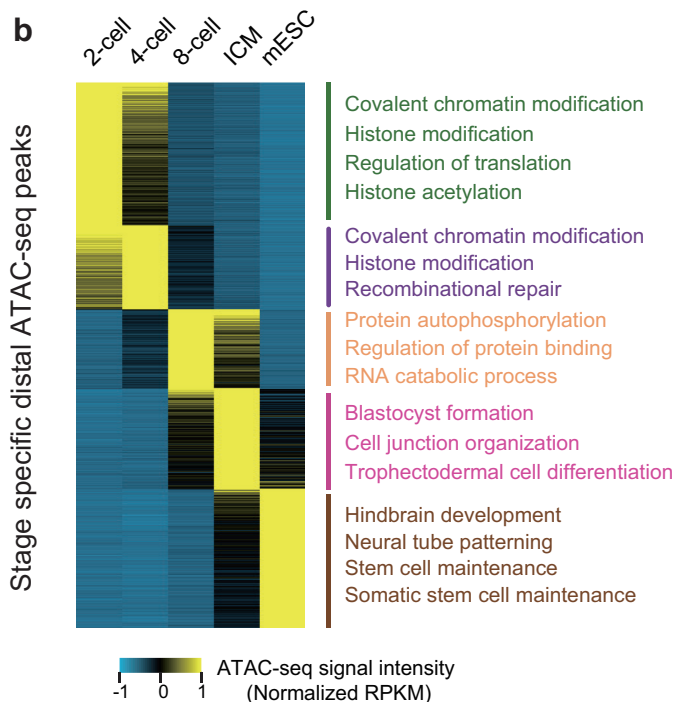
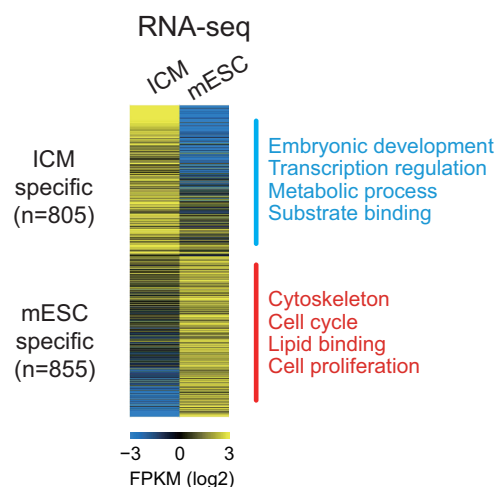


Extended Data Figure 6 | Widespread presence of repetitive elements in accessible chromatin in early mouse embryos. **a**, Bar charts showing the percentages of distal ATAC-seq peaks in early embryos that overlap with enhancers in somatic tissues identified previously²⁰. **b**, Bar charts showing the percentages of the RNA-seq reads mapped to RepBase (repeat reads) compared to all mappable reads (repeat reads + single-copy gene reads). The error bars denote the standard deviations of two biological replicates of RNA-seq for each stage. **c**, Bar charts showing the number of peaks with no overlap with repeats, with any regions in the peak overlapping with repeats, or with at least 20% of the peak regions overlapping with repeats.

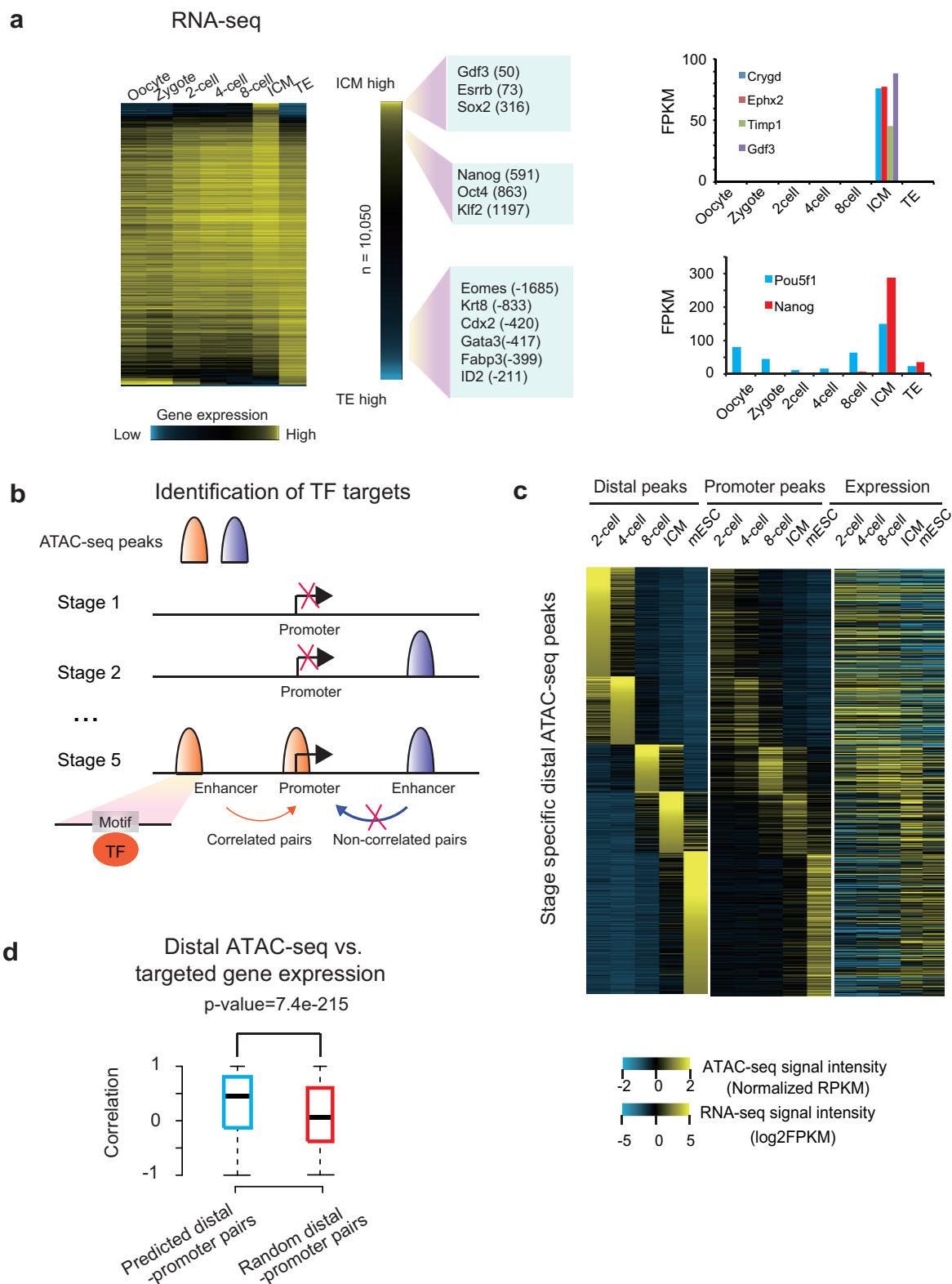
d, The snapshots of promoter (left) or distal (middle and right) ATAC-seq peaks near various types of repeats. **e**, Percentages of stage-specific genes (ZGA-only) of which the TSSs directly overlap with repeats are shown as bar graphs. A similar analysis result for random TSSs overlapping with repeats is included as control. **f**, The percentages of different classes of repeats that overlap with TSSs of stage-specific genes are shown. **g**, Heat maps show the enrichment of repeats in promoter and distal H3K27ac ChIP-seq peaks of early embryos and somatic tissues. **h**, The densities of various families of repeats near genes are shown.

a Stage specificity of ATAC-seq peaks**c** Chromatin regulator gene expression

Extended Data Figure 7 | Analysis of stage-specific distal ATAC-seq peaks. **a**, Histogram shows the distribution of the Shannon entropy of enrichment levels for ATAC-seq promoter peaks and distal peaks across early embryos and mESCs. **b**, A heat map showing the ATAC-seq enrichment at stage specific distal ATAC-seq peaks (left). The functional enrichment of genes nearby (by the GREAT analysis²⁵) is also shown (right). **c**, Box plots showing the expression levels of all chromatin regulator genes (as identified by the GREAT tool) near stage-specific

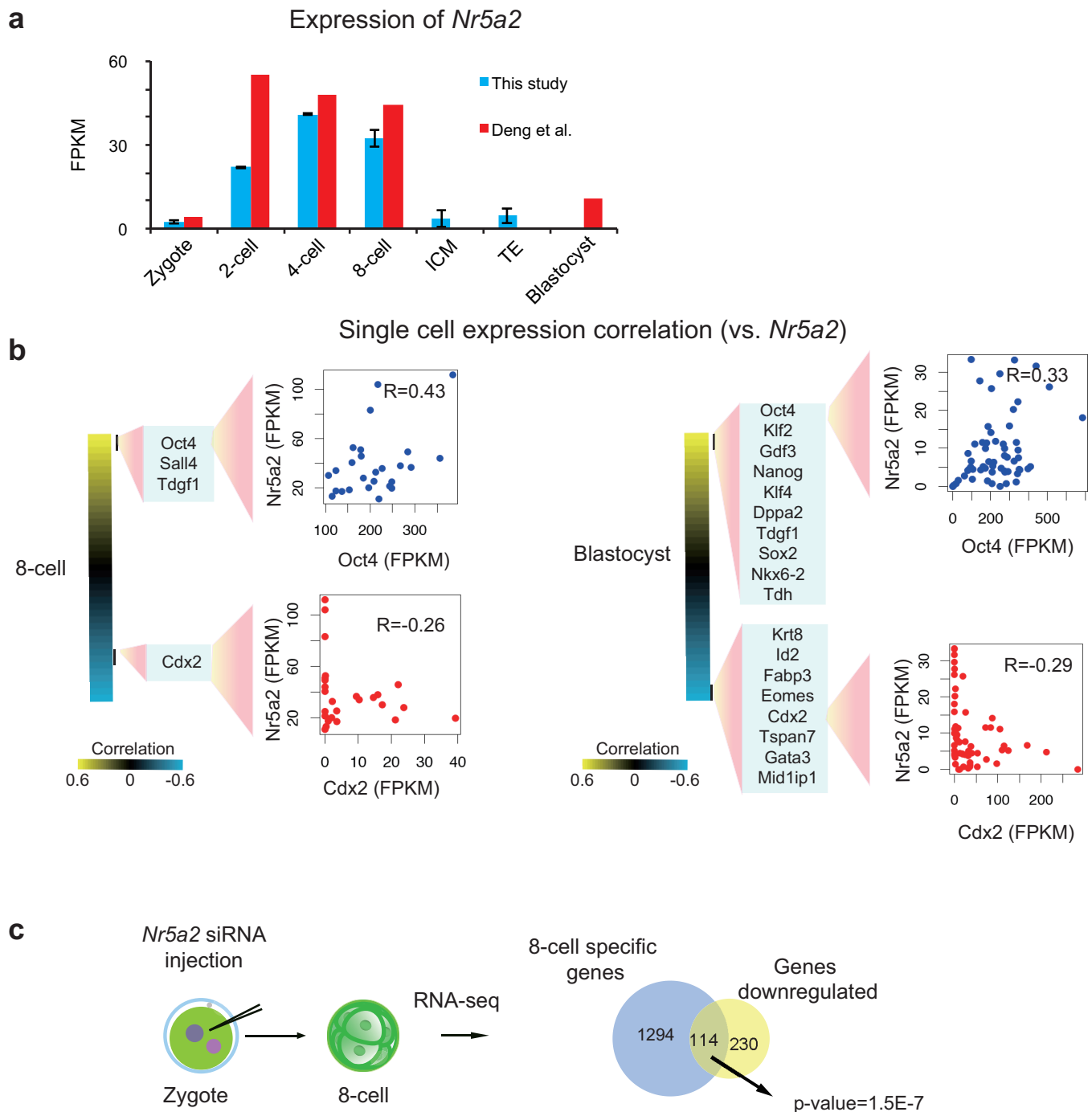
**e**

distal ATAC-seq peaks in early development. **d**, Scatter plots comparing the gene expression changes between ICMs and mESCs (log ratio of FPKM) for each gene in this study and in Tang *et al.*⁴⁶. Genes that are lowly expressed in both ICMs and mESCs (FPKM < 1) were removed. **e**, A heat map showing the expression levels of the ICM- and mESC-specific genes (identified in both this study and Tang *et al.*⁴⁶) (left). The Gene Ontology terms are also shown (right).



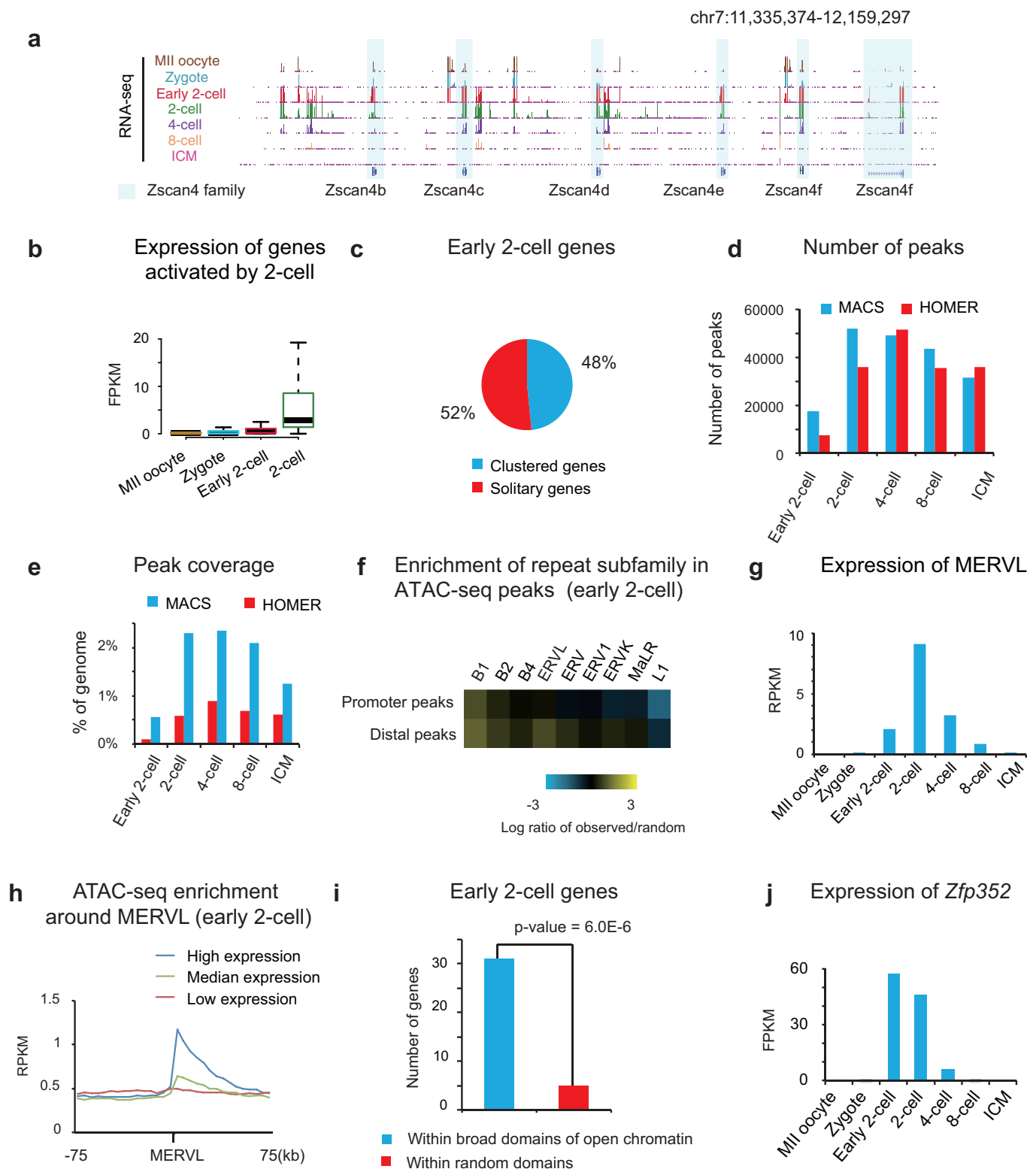
Extended Data Figure 8 | Analysis of single-cell RNA-seq and identification of putative promoter-enhancer pairs in early embryos. a, A heat map shows the differential expression levels between ICM and TE for all genes ranked by their ICM:TE expression ratio (left). Genes that with low expression in both ICM and TE (FPKM <1) were removed. Selected ICM and TE regulators/markers are shown with their differential expression rank indicated in parentheses (middle). Positive and negative ranks indicate the ranking in ICM-specific and TE-specific genes, respectively. Examples of ICM-specific genes are shown on the right, including those exclusively expressed in ICM (right top) or known ICM regulators (right bottom). **b,** Schematic shows the method to identify target genes of a TF. The correlation of ATAC-seq enrichment of distal peaks and their nearby (within 100 kb)

promoter peaks was calculated across all developmental stages. Distal-promoter peak pairs showing high correlation are considered as putative enhancer-promoter pairs (see Supplementary Information). Candidate regulatory TFs were identified by searching motifs in distal peaks, and distal-peak linked promoters were defined as the potential targets of TFs. **c,** Heat maps showing the ATAC-seq enrichment at stage-specific distal ATAC-seq peaks, their predicted target promoters, and the expression of corresponding genes. Only genes that are not expressed in oocytes (ZGA-only) were included in this analysis. **d,** Box plots showing the correlation between enrichment levels of distal ATAC-seq peaks and expression of their predicted target genes, compared to random distal peak-promoter peak pairs. A *P* value based on the *t*-test is shown.



Extended Data Figure 9 | NR5A2 regulates expression of key TFs for ICM and TE at the 8-cell stage. **a**, Bar charts showing the expression levels of *Nr5a2* in preimplantation embryos using data from this study or Deng *et al.*¹². The error bars denote the standard deviation of FPKM values across two RNA-seq replicates. **b**, Heat maps show the correlation between the expression of *Nr5a2* and all genes across individual single cells from

the 8-cell embryos (left) or blastocysts (right) using data from Deng *et al.*¹². Scatter plots comparing expression of *Nr5a2* with *Pou5f1* or *Cdx2* across individual cells from the 8-cell embryos or blastocysts. **c**, Schematic of the *Nr5a2* knockdown experiments. A Venn diagram shows the overlap between *Nr5a2*-knockdown downregulated genes and 8-cell-specific genes. A *P* value based on the hypergeometric distribution is shown.



Extended Data Figure 10 | Transcription and chromatin states in minor ZGA.

a, A snapshot of the UCSC genome browser shows an example of early 2-cell gene family (*Zscan4*) that reside in clusters in the genome. **b**, Box plots show the expression levels of ZGA-only genes that are activated by the 2-cell stage (activated either at zygote, early 2-cell or 2-cell stages). **c**, Pie chart shows the percentages of early 2-cell genes that reside in clusters or are solitary. **d**, Bar chart shows the numbers of ATAC-seq peaks identified at each stage. Two programs were used (MACS and HOMER²⁶) to verify the ATAC-seq peak analyses in early 2-cell stages. **e**, Bar chart shows the genome coverages by ATAC-seq peaks at different stages called either by MACS or HOMER. **f**, The enrichment (log ratio of observed to random) of repeat subfamily in ATAC-seq promoter and

distal peaks in the early 2-cell embryos. **g**, Bar chart shows the expression levels of MERV in oocyte and preimplantation embryos. **h**, The average ATAC-seq enrichment in the early 2-cell embryos is shown for regions near MERVs with high, median and low levels of expression. The MERV region is not shown due to the low mappability. **i**, Bar chart shows the numbers of early 2-cell genes that fall into the broad domains of open chromatin (of equal length as each corresponding broad domain) was performed as a control. The *P* value was calculated based on the hypergeometric distribution. **j**, Bar chart shows the expression levels of *Zfp352* in oocyte and preimplantation embryos.

A Neptune-sized transiting planet closely orbiting a 5–10-million-year-old star

Trevor J. David¹, Lynne A. Hillenbrand¹, Erik A. Petigura², John M. Carpenter³, Ian J. M. Crossfield⁴, Sasha Hinkley⁵, David R. Ciardi⁶, Andrew W. Howard⁷, Howard T. Isaacson⁸, Ann Marie Cody⁹, Joshua E. Schlieder⁹, Charles A. Beichman⁶ & Scott A. Barenfeld¹

Theories of the formation and early evolution of planetary systems postulate that planets are born in circumstellar disks, and undergo radial migration during and after dissipation of the dust and gas disk from which they formed^{1,2}. The precise ages of meteorites indicate that planetesimals—the building blocks of planets—are produced within the first million years of a star's life³. Fully formed planets are frequently detected on short orbital periods around mature stars. Some theories suggest that the *in situ* formation of planets close to their host stars is unlikely and that the existence of such planets is therefore evidence of large-scale migration^{4,5}. Other theories posit that planet assembly at small orbital separations may be common^{6–8}. Here we report a newly born, transiting planet orbiting its star with a period of 5.4 days. The planet is 50 per cent larger than Neptune, and its mass is less than 3.6 times that of Jupiter (at 99.7 per cent confidence), with a true mass likely to be similar to that of Neptune. The star is 5–10 million years old and has a tenuous dust disk extending outward from about twice the Earth–Sun separation, in addition to the fully formed planet located at less than one-twentieth of the Earth–Sun separation.

The star [PGZ2001] J161014.7–191909, hereafter K2-33, is an M-type star several million years (Myr) old that was observed by NASA's Kepler Space Telescope during campaign 2 of the K2 mission. The star was identified as one of more than 200 candidate planet hosts in a systematic search for transits in K2 data⁹. As part of our ongoing study of the pre-main-sequence population of Upper Scorpius observed by K2, we independently verified and analysed the planetary transit signal. We acquired radial velocity and high spatial resolution observations at the W. M. Keck Observatory to confirm the detection of the planet, named K2-33 b, and to measure its size and mass.

Within the 77.5-day photometric time series of K2-33 ($K_p = 14.3$ mag), there are periodic dimmings of 0.23% lasting 4.2 h and occurring every 5.4 days (Fig. 1). The ensemble of transits are detected at a combined signal-to-noise ratio of about 32. During the K2 observations, cool, dark regions on the stellar surface (starspots) rotated in and out of view, producing semi-sinusoidal brightness variations of ~3% peak-to-trough amplitude with a periodicity of 6.3 ± 0.2 days (Extended Data Fig. 1). We removed the starspot variability before modelling the transit events. We fitted the transit profiles using established methods¹⁰, measuring the planet's size relative to its host star and its orbital geometry (Table 1).

K2-33 is a member of the Upper Scorpius OB association^{11,12}, the nearest site to Earth of recent massive star formation (at a distance of 145 ± 20 pc). Approximately 20% of low-mass stars in Upper Scorpius host protoplanetary disks¹³, indicating that planet formation is ongoing in the region but in an advanced stage or completed for the majority of stars. The age of the stellar association is 5–10 Myr,

as assessed from kinematics, the Hertzsprung–Russell diagram, and eclipsing binary analyses. The youth of K2-33 itself is based on the spectroscopic indicators of enhanced hydrogen emission and lithium absorption^{11,12}, which we confirm from Keck spectra (Table 1). Furthermore, the stellar rotation rate we measure via broadening of absorption lines in the spectra and via the starspot period (Table 1), is rapid relative to field-age stars of similar mass¹⁴. We determined the star's systemic radial velocity (Table 1) to be consistent with the mean value for Upper Scorpius members¹⁵. Previously, proper motions were used to assess the probability of membership in the association at 99.9% (ref. 16). Finally, the positions of the star in the Hertzsprung–Russell and temperature–density diagrams (Extended Data Fig. 2) are consistent with the sequence of low-mass members of Upper Scorpius¹².

The inferred planetary size and mass depend directly upon the host star size and mass. We evaluated the effective temperature and luminosity from our newly determined spectral type (Table 1), extinction-corrected catalogue near-infrared photometry, and empirical pre-main-sequence calibrations^{17,18}. With the temperature and luminosity, we derive a stellar radius from the Stefan–Boltzmann law of $R_* = (1.1 \pm 0.1)R_\odot$, where R_\odot is the radius of the Sun. The radius uncertainty is calculated accounting for recommended errors in temperature¹⁷, photometric errors, and assuming an association depth comparable to its width on the sky. Combining the stellar radius with the planet-to-star radius ratio determined from the K2 light-curve fit, we infer a planetary radius for the companion of $R_p = (5.8 \pm 0.6)R_\oplus$, where R_\oplus is Earth's radius, or about 50% larger than Neptune.

We estimate a stellar mass of $M_* = (0.31 \pm 0.05)M_\odot$, where M_\odot is the mass of the Sun, by interpolation among pre-main-sequence stellar evolution models¹⁹, consistent with a previously reported value²⁰. The mass uncertainty assumes normal error distributions in temperature and luminosity. As there is no evidence for radial velocity variations among four high-dispersion Keck spectra (Extended Data Table 1), the planet mass is constrained from the maximum-amplitude Keplerian curve that is consistent within the errors with all radial velocity measurements (Extended Data Fig. 3). Given the transit ephemeris and assuming a circular, edge-on orbit, the expected stellar reflex velocity is a sinusoid having a single free parameter: the semi-amplitude. Radial velocity semi-amplitudes of $>900 \text{ m s}^{-1}$ are ruled out at 99.7% confidence, corresponding to a 3σ upper limit on the mass of K2-33 b of $3.6M_{\text{Jup}}$, where M_{Jup} is the mass of Jupiter.

The true mass of K2-33 b is likely to be at least an order of magnitude smaller. There are seven known exoplanets of similar size ($R_p = 4.8R_\oplus$ – $6.6R_\oplus$) with densities measured to 50% or better. These planets have masses ranging from $(6.3 \pm 0.8)M_\oplus$ (for Kepler-87 c; where M_\oplus is Earth's mass) to $(69 \pm 11)M_\oplus$ (for CoRoT-8 b) owing to varying core masses. Thus, plausible masses of K2-33 b range from about $6M_\oplus$ to

¹Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, California 91125, USA. ²Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. ³Joint ALMA Observatory, Avenida Alonso de Córdova 3107, Vitacura, Santiago, Chile. ⁴Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. ⁵Physics Department, University of Exeter, Stocker Road, Exeter EX4 4QL, UK. ⁶NASA Exoplanet Science Institute, California Institute of Technology, Pasadena, California 91125, USA. ⁷Institute for Astronomy, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, USA. ⁸Department of Astronomy, University of California, Berkeley, California 94720, USA. ⁹NASA Ames Research Center, Mountain View, California 94035, USA.

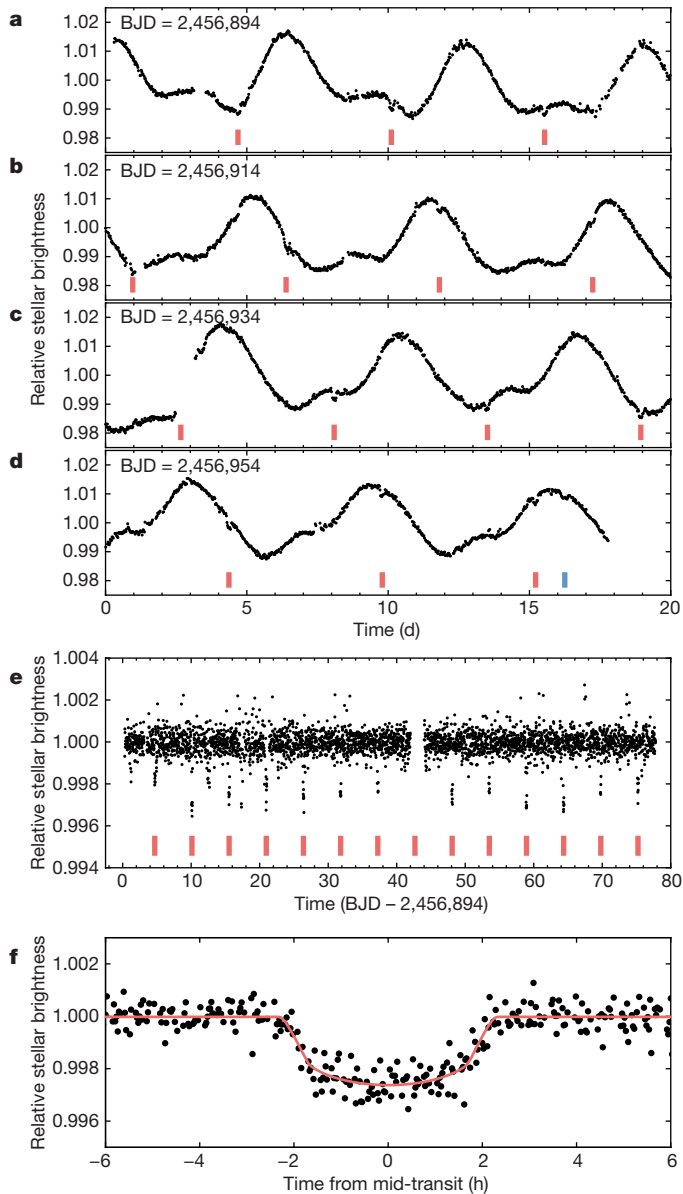


Figure 1 | Light curve of K2-33. a–d, K2 photometry in twenty-day segments. K2-33 varies in brightness by $\sim 3\%$ every 6.3 days owing to the rotation of its spotted surface. The planet K2-33 b transits its star every 5.4 days (red ticks). Another potential transit (blue tick in d), distinct from the K2-33 b transits, is possibly caused by a second planet with an orbital period of > 77.5 days. e, Stellar variability was removed before fitting transits. The data gap is due to excluded observations where the variability fit inadequately captured a systematic artefact in the light curve. f, K2 photometry folded on the planet's orbital period with a transit model fit (red curve). BJD, barycentric Julian date.

$70M_{\oplus}$, corresponding to radial velocity semi-amplitudes of $5\text{--}56\text{ m s}^{-1}$. An even lower mass may be implied if the young planet is still undergoing Kelvin–Helmholtz radial contraction.

A semi-major axis of 0.04 au ($\sim 8R_*$) is measured for K2-33 b from the orbital period and Kepler's third law, adopting the value of M_* in Table 1. The orbit is near the silicate dust sublimation radius, as well as the co-rotation radius, where some protoplanetary disk theories predict a magnetospheric cavity extending to the stellar surface²¹. At this separation, the blackbody equilibrium temperature of the planet is 850 K .

We have interpreted the observed transit as a single planet orbiting K2-33. Other interpretations involving eclipsing stellar binaries residing within the K2 aperture (Extended Data Fig. 4) would be diluted by the light of K2-33 and could potentially mimic the observed transit.

Table 1 | System properties of K2-33

Parameter	Value	Uncertainty
Stellar properties		
2MASS designation	J16101473–1919095	
EPIC designation	205117205	
Right ascension (J2000.0)	16 h 10 min 14.738 s	
Declination (J2000.0)	$-19^\circ 19' 09.55''$	
Proper motion, RA, μ_α (mas yr $^{-1}$)	-9.8	± 1.7
Proper motion, dec., μ_δ (mas yr $^{-1}$)	-24.2	± 1.8
Kepler magnitude, Kp (mag)	14.3	
Cluster distance, d (pc)	145	± 20
Kinematic distance, d_{kin} (pc)	139	± 11
Spectral type	M3	± 0.5
V-band extinction, A_V (mag)	1.29	
Luminosity, $\log(L_*/L_\odot)$ (dex)	-0.83	± 0.07
Effective temperature, T_{eff} (K)	3,410	± 75
Stellar radius, R_* (R_\odot)	1.1	± 0.1
Stellar mass, M_* (M_\odot)	0.31	± 0.05
Mean stellar density, ρ_* (g cm $^{-3}$)	0.34	± 0.12
Surface gravity, $\log(g_*)$ (dex)	3.84	± 0.16
Stellar rotation period, P_{rot} (days)	6.3	± 0.2
Systemic radial velocity, γ (km s $^{-1}$)	-6.6	± 0.1
Projected rotational velocity, $v \sin(i)$ (km s $^{-1}$)	5–9	
H α equivalent width (\AA)	-1.3	± 0.1
H β equivalent width (\AA)	-1.05	± 0.05
Li I 6,708 \AA equivalent width (\AA)	0.60	± 0.05
Light-curve modelling parameters		
Orbital period (days)	5.42513	$+0.00028$ -0.00029
Time of mid-transit, t_0 (BJD; days)	2,456,936.6665	$+0.0012$ -0.0012
Transit duration, T_{14} (hours)	4.22	$+0.15$ -0.10
Planet-to-star radius ratio, R_p/R_*	0.0476	$+0.0035$ -0.0017
Scaled semi-major axis, R_*/a	0.109	$+0.033$ -0.012
Impact parameter, b	0.49	$+0.26$ -0.33
Inclination, i (deg)	86.9	$+2.2$ -3.1
Mean stellar density, $\rho_{*,\text{circ}}$ (g cm $^{-3}$)	0.49	$+0.21$ -0.27
Linear limb darkening coefficient, u	0.603	$+0.052$ -0.053
Planet properties		
Planet radius, R_p (R_\oplus)	5.76	$+0.62$ -0.58
Planet mass, M_p (M_{Jup})	< 3.6	
Semi-major axis, a (AU)	0.0409	$+0.0021$ -0.0023
Blackbody equilibrium temperature, T_{eq} (K)	850	± 50

Right ascension and declination originate from 2MASS, proper motions from UCAC4, and Kepler magnitude from the Ecliptic Plane Input Catalog. The mean cluster distance³⁰ is assumed, with uncertainty equal to the presumed cluster depth¹². Quoted transit parameters and uncertainties are the medians and 15.87%, 84.13% percentiles of the posterior distributions.

Such a putative eclipsing stellar binary could be associated with (that is, gravitationally bound to) K2-33, or unassociated but aligned by chance. Given constraints from our suite of follow-up observations, we show that the chance of an eclipsing stellar binary producing the observed transit is vanishingly small.

We first consider unassociated eclipsing stellar binaries, which we characterize by their sky-projected separation from K2-33 and their brightness relative to K2-33 in the Kepler bandpass, ΔKp . Eclipse

depths may not exceed 100%; thus eclipsing stellar binaries with $\Delta Kp > 6.6$ mag cannot account for the 0.23% observed transit depth. We do not detect closely projected companions in seeing-limited and multi-epoch adaptive optics images (Extended Data Fig. 5), nor in searches for secondary lines in high-resolution optical spectra. These observational constraints, shown in Fig. 2a, eliminate nearly all scenarios involving unassociated eclipsing stellar binaries. The probability of an eclipsing stellar binary lurking in the remaining parameter space is $\ll 4 \times 10^{-6}$ (see Methods).

We now consider associated eclipsing stellar binaries in terms of their physical distance to K2-33, d , and ΔKp . As in the case of unassociated eclipsing stellar binaries, our imaging and spectroscopic data eliminate the vast majority of associated eclipsing stellar binary configurations. Additionally, the lack of detectable line-of-sight acceleration over the baseline of the observations rules out associated eclipsing stellar binaries in a mass-dependent manner (see Methods). The constraints provided by these complementary observations are depicted in Fig. 2b. However, some scenarios having $d = 1\text{--}3$ AU and $\Delta Kp = 2\text{--}6$ mag cannot be conclusively eliminated. Nearly all of these scenarios involve a planet orbiting either K2-33 or an undetected companion. If orbiting K2-33, the planet radius is at most 7.6% larger than the value reported in Table 1 (that is, within quoted uncertainties) owing to dilution from a companion. If orbiting a stellar or substellar companion to K2-33, the planet radius is at most $\sim 1.8R_{\text{Jup}}$, where M_{Jup} is the mass of Jupiter. Only for $\Delta Kp \geq 6$ mag does the radius of the transiting object correspond to a mass of $\geq 13M_{\text{Jup}}$ at 5–10 Myr. However, coeval eclipsing brown dwarfs probably would not produce eclipse depths $> 50\%$, and thus contrasts of $\Delta Kp \geq 5.8$ mag need not be considered. Furthermore, the low occurrence of brown dwarfs²² combined with the lack of observed secondary eclipses make such configurations extremely unlikely. Given the Kepler photometry and observational constraints, we quantitatively assessed the overall false-positive probability, using an established statistical framework²³ (see Methods). We found that scenarios involving a single star and planet are 10^{11} times more likely than scenarios involving eclipsing stellar binaries.

Spitzer Space Telescope observations of K2-33 revealed 24- μm emission in excess of the expected stellar photosphere by 50%, indicating the presence of cool circumstellar dust²⁴. There is an absence, however, of warm dust close to the star, given the lack of similar infrared excess at wavelengths shorter than 16 μm (ref. 25). The spectral energy distribution is best fitted by including a dust component at 122 K having an inner edge at 2.0 AU. These data suggest that the inner regions of the previously present protoplanetary disk have cleared. Supporting this inference is the modest H α signature (Table 1), which is consistent with chromospheric emission and indicates that the star is not accreting gas.

At the age of K2-33, it is unclear whether the dust structure consists of debris resulting from the collisional grinding of planetesimals, or whether it is a remnant of the initial dust- and gas-rich disk. One possibility is that the inner-disk regions have been cleared of dust by the gravitational influence of one or more planetary mass bodies²⁶. Our detection of a short-period planet in a transitional disk lends support to this explanation.

A flux upper limit at 880 μm from the Atacama Large Millimeter Array combined with the measured Spitzer fluxes yields a constraint on the mass of dust remaining in the disk of less than $0.2M_{\oplus}$ (ref. 27). Additionally, CO emission, a tracer of molecular hydrogen, was not detected²⁷, indicating that the primordial gas disk also has largely or entirely dissipated.

The transiting planet around the young star K2-33 provides direct evidence that large planets can be found at small orbital separations shortly after dispersal of the nebular gas. Migration via tidal circularization of an eccentric planet, through, for example, the Kozai–Lidov mechanism, planet–planet scattering, or secular chaos, proceeds over timescales much greater than disk dispersal timescales, and thus cannot explain the planet's current orbit. *In situ* formation, or formation

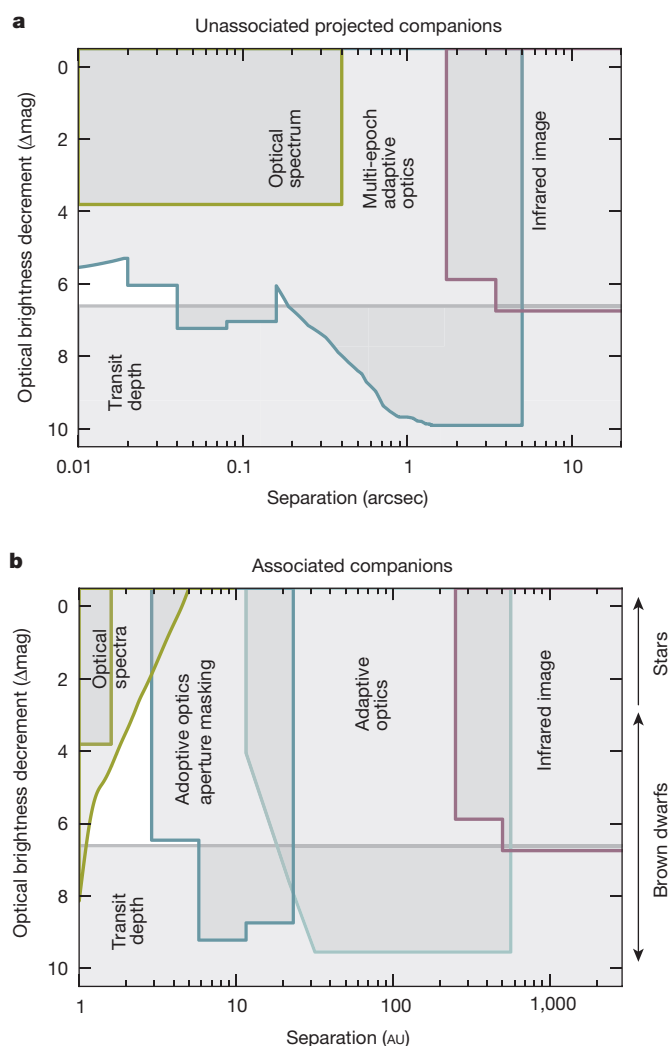


Figure 2 | Constraints on astrophysical false-positive scenarios.

To confirm the planetary nature of K2-33 b, we considered and eliminated nearly all false-positive scenarios involving eclipsing stellar binaries. **a**, The domain of sky-projected separation and contrast of a putative unassociated eclipsing stellar binary, aligned with K2-33 by chance. The blue outlined region shows eclipsing stellar binaries eliminated using multi-epoch adaptive optics imaging, which leverages stellar proper motion to provide sensitivity at all separations within 5". Green and purple regions represent constraints from optical spectra and seeing-limited imaging, respectively. Finally, eclipsing stellar binaries below the grey line cannot account for the observed transit depth and are eliminated. **b**, Limits on eclipsing stellar binaries associated with (that is, gravitationally bound to) K2-33. Constraints from imaging and spectroscopy are shown as a function of physical separation. The lack of detectable stellar acceleration provides an additional diagonal constraint at top left. Δmag , difference in magnitude.

at a larger separation followed by migration within the gas disk, are permitted scenarios given current observations.

Interestingly, large planets are rarely found close to mature low-mass stars; fewer than 1% of M-dwarfs host Neptune-sized planets with orbital periods of < 10 days²⁸, while $\sim 20\%$ host Earth-sized planets in the same period range²⁹. This may be a hint that K2-33 b is still contracting, losing atmosphere, or undergoing radial migration. Future observations may test these hypotheses, and potentially reveal where in the protoplanetary disk the planet formed.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 March; accepted 22 April 2016.

Published online 20 June 2016.

1. Kuiper, G. P. On the origin of the Solar System. *Proc. Natl Acad. Sci. USA* **37**, 1–14 (1951).
2. Kley, W. & Nelson, R. P. Planet-disk interaction and orbital evolution. *Annu. Rev. Astron. Astrophys.* **50**, 211–249 (2012).
3. Connelly, J. N. *et al.* The absolute chronology and thermal processing of solids in the solar protoplanetary disk. *Science* **338**, 651–655 (2012).
4. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380**, 606–607 (1996).
5. Rafikov, R. R. Atmospheres of protoplanetary cores: critical mass for nucleated instability. *Astrophys. J.* **648**, 666–682 (2006).
6. Hansen, B. M. S. & Murray, N. Migration then assembly: formation of Neptune-mass planets inside 1 AU. *Astrophys. J.* **751**, 158 (2012).
7. Chatterjee, S. & Tan, J. C. Inside-out planet formation. *Astrophys. J.* **780**, 53 (2014).
8. Boley, A. C., Granados Contreras, A. P. & Gladman, B. The *in situ* formation of giant planets at short orbital periods. *Astrophys. J.* **817**, L17 (2016).
9. Vanderburg, A. *et al.* Planetary candidates from the first year of the K2 mission. *Astrophys. J. Suppl. Ser.* **222**, 14 (2016).
10. Crossfield, I. J. M. *et al.* A nearby M star with three transiting super-Earths discovered by K2. *Astrophys. J.* **804**, 10 (2015).
11. Preibisch, T., Guenther, E. & Zinnecker, H. A large spectroscopic survey for young low-mass members of the Upper Scorpius OB association. *Astron. J.* **121**, 1040–1049 (2001).
12. Preibisch, T., Brown, A. G. A., Bridges, T., Guenther, E. & Zinnecker, H. Exploring the full stellar population of the Upper Scorpius OB association. *Astron. J.* **124**, 404–416 (2002).
13. Luhman, K. L. & Mamajek, E. E. The disk population of the Upper Scorpius association. *Astrophys. J.* **758**, 31 (2012).
14. McQuillan, A., Aigrain, S. & Mazeh, T. Measuring the rotation period distribution of field M dwarfs with Kepler. *Mon. Not. R. Astron. Soc.* **432**, 1203–1216 (2013).
15. Kurosawa, R., Harries, T. J. & Littlefair, S. P. Radial and rotational velocities of young brown dwarfs and very low-mass stars in the Upper Scorpius OB association and the ρ Ophiuchi cloud core. *Mon. Not. R. Astron. Soc.* **372**, 1879–1887 (2006).
16. Bouy, H. & Martín, E. L. Proper motions of cool and ultracool candidate members in the Upper Scorpius OB association. *Astron. Astrophys.* **504**, 981–990 (2009).
17. Herczeg, G. J. & Hillenbrand, L. A. Empirical isochrones for low mass stars in nearby young associations. *Astrophys. J.* **808**, 23 (2015).
18. Peca, M. J. & Mamajek, E. E. Intrinsic colors, temperatures, and bolometric corrections of pre-main-sequence stars. *Astrophys. J. Suppl. Ser.* **208**, 9 (2013).
19. Baraffe, I., Homeier, D., Allard, F. & Chabrier, G. New evolutionary models for pre-main sequence and main sequence low-mass stars down to the hydrogen-burning limit. *Astron. Astrophys.* **577**, A42 (2015).
20. Kraus, A. L. & Hillenbrand, L. A. The role of mass and environment in multiple-star formation: a 2MASS survey of wide multiplicity in three young associations. *Astrophys. J.* **662**, 413–430 (2007).
21. Armitage, P. J. Dynamics of protoplanetary disks. *Annu. Rev. Astron. Astrophys.* **49**, 195–236 (2011).
22. Bowler, B. P., Liu, M. C., Shkolnik, E. L. & Tamura, M. Planets around low-mass stars (PALMS). IV. The outer architecture of M dwarf planetary systems. *Astrophys. J. Suppl. Ser.* **216**, 7 (2014).
23. Morton, T. D. VESPA: false positive probabilities calculator. *Astrophys. Source Code Library* ascl:1503.011, <http://ascl.net/1503.011> (2015).
24. Carpenter, J. M., Mamajek, E. E., Hillenbrand, L. A. & Meyer, M. R. Debris disks in the Upper Scorpius OB association. *Astrophys. J.* **705**, 1646–1671 (2009).
25. Carpenter, J. M., Mamajek, E. E., Hillenbrand, L. A. & Meyer, M. R. Evidence for mass-dependent circumstellar disk evolution in the 5 Myr old Upper Scorpius OB association. *Astrophys. J.* **651**, L49–L52 (2006).
26. Trilling, D. E. *et al.* Debris disks around Sun-like stars. *Astrophys. J.* **674**, 1086–1105 (2008).
27. Barenfeld, S., Carpenter, J. M., Ricci, L. & Isella, A. ALMA observations of circumstellar disks in the Upper Scorpius OB association. *Astrophys. J.* (in the press); preprint at <https://arxiv.org/abs/1605.05772>.
28. Dressing, C. D. & Charbonneau, D. The occurrence rate of small planets around small stars. *Astrophys. J.* **767**, 95 (2013).
29. Dressing, C. D. & Charbonneau, D. The occurrence of potentially habitable planets orbiting M dwarfs estimated from the full Kepler dataset and an empirical measurement of the detection sensitivity. *Astrophys. J.* **807**, 45 (2015).
30. de Zeeuw, P. T., Hoogerwerf, R., de Bruijne, J. H. J., Brown, A. G. A. & Blaauw, A. A HIPPARCOS census of the nearby OB associations. *Astron. J.* **117**, 354–399 (1999).

Acknowledgements We thank S. Metchev, K. Batygin, B. Benneke, K. Deck, J. Fuller and A. Shporer for discussions, M. Ireland for software used in the aperture masking analysis, and A. Kraus for contributing to the 2011 Keck/NIRC2 data acquisition. T.J.D. is supported by an NSF Graduate Research Fellowship under Grant DGE1144469. E.A.P. is supported through a Hubble Fellowship. I.J.M.C. is supported through a Sagan Fellowship. A.W.H. acknowledges funding from NASA grant NNX16AE75G and NASA Research Support Agreement 1541779. This paper includes data collected by the Kepler mission, funded by the NASA Science Mission directorate. Some data presented here were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. We acknowledge the important cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community and we are fortunate to be able to conduct observations from this mountain.

Author Contributions T.J.D. noted the object as a young star, prepared the light curve, validated the transit, and led the overall analysis and the writing of the paper. L.A.H. analysed the 2015 and 2016 Keck/HIRES spectra, participated in team organization, performed general analysis, and contributed substantially to the writing of the paper. E.A.P. analysed raw K2 photometry and validated the transit, provided general guidance on exoplanets and false positives, and contributed substantially to the writing of the paper. J.M.C. was involved in the K2 proposal that included the object, and was principal investigator on the 2011 Keck/NIRC2, 2015 Keck/HIRES, and ALMA observations of the object. I.J.M.C. led the transit fitting and VESPA analysis. A.M.C. analysed the raw K2 photometry and validated the transit. A.W.H. and H.T.I. obtained, reduced, and analysed the 2016 Keck/HIRES spectra. D.R.C. led the clear aperture adaptive optics contrast curve analysis for the 2016 and 2011 data. C.A.B. wrote the proposal for 2016 Keck/NIRC2 follow-up of K2 sources and participated in the observations of the object. S.H. analysed the Keck/NIRC2 aperture masking data and assessed the temperature of the circumstellar dust. J.E.S. provided a rotational velocity analysis and calculated the kinematic distance. S.A.B. took the 2015 Keck/HIRES spectrum and analysed the ALMA data (referred to, but published separately in ref. 27).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.J.D. (tjd@astro.caltech.edu).

Reviewer Information Nature thanks A. Collier Cameron and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Stellar membership and properties. The partial kinematics of K2-33 (Table 1) can be combined with the galactic velocity of the Upper Scorpius subgroup to estimate distance and predict radial velocity. We calculated these parameters following established methods³¹, adopting the mean UVW galactic velocities of the subgroup³², and estimated uncertainties using Monte Carlo sampling. We find $d_{\text{kin}} = 139 \pm 11$ pc, consistent with the mean subgroup distance from Earth 145 ± 2 pc (ref. 30), and radial velocity -7.3 ± 0.5 km s⁻¹, within 2σ of the systemic radial velocity we measure from multiple Keck/HIRES spectra (Extended Data Table 1).

We determined spectral type from the high-dispersion spectra and adopted an empirical spectral type to temperature conversion calibrated for young stars^{17,18} to estimate effective temperature (Table 1). From 2MASS photometry and the appropriate intrinsic $J-H$ colour for the spectral type¹⁸, we calculated the $J-H$ colour excess, $E(J-H) = 0.10$ mag. Assuming an extinction law we found visual and J-band extinctions of $A_V = 1.29$ mag and $A_J = 0.30$ mag, respectively. After correcting for extinction, we used the appropriate empirical J-band bolometric correction for the spectral type^{17,18} and a distribution of distances to calculate luminosity. While the mean association distance is known precisely, its large sky-projected area (about 150 deg²) suggests that the association depth is substantial. Results from a secular parallax study indicate an association distance spread of <50 pc, with consideration of the angular diameter corresponding to a 35-pc spread^{12,33} at the mean distance. In calculating luminosity, we conservatively considered a uniform distribution of distances in a cubic volume 40 pc on each side, centred on the mean association distance of 145 pc (ref. 30). Uncertainty is calculated from Monte Carlo sampling, accounting for photometric errors, recommended errors in temperature¹⁷, and distance uncertainties. We propagated luminosity and temperature errors through in calculating the radius uncertainties using the Stefan–Boltzmann law.

Detailed modelling of the transit profile provides a constraint on mean stellar density, assuming a circular orbit³⁴. We found a mean stellar density of $\rho_{*,\text{circ}} = 0.49$ g cm⁻³, consistent with the value implied by our adopted M_* and R_* (Table 1). From the posterior distribution of mean stellar densities and a normal distribution in effective temperature, we interpolated between pre-main-sequence models¹⁹ to determine a stellar mass and age of $M_* = (0.30 \pm 0.04)M_\odot$ and $t = 5^{+1}_{-2}$ Myr, respectively (Extended Data Fig. 2). However, we conservatively adopt the ensemble age of the association, as we consider it more robust than the age of an individual star.

Stellar rotation and independent assessment of the stellar radius. We measured the stellar rotation period as $P_{\text{rot}} = 6.3 \pm 0.2$ days from a Lomb–Scargle periodogram^{35,36} of the light curve. Uncertainty was determined from the half-width at half-maximum of a Gaussian fit to the periodogram peak. Extended Data Fig. 1 depicts the light curve folded on the rotational period.

The stellar rotation speed as projected along the line of sight, $v \sin i_*$, was estimated from the spectra by artificially broadening an absorption spectrum of a slowly rotating stellar template of similar temperature to K2-33, acquired using the same spectrograph and set-up. The range of plausible rotational velocities is constrained through minimization of the residuals between the broadened template and the observed spectrum. We find a most likely projected rotation velocity $v \sin i_* \approx 5\text{--}9$ km s⁻¹. Combined with the rotation period measured from the light curve, we used the projected rotational velocity to determine an independent estimate of the stellar radius modulated by the sine of the stellar inclination of $R_* \sin i_* = v \sin i_* \cdot P_{\text{rot}} / 2\pi = (0.85 \pm 0.25)R_\odot$, where we quote a 95% confidence interval assuming a uniform distribution in $v \sin i_*$.

Our value for $R_* \sin i_*$ is consistent with the Stefan–Boltzmann radius within 2σ . Two effects could bias $R_* \sin i_*$ away from the true value of R_* : (1) the surface features dominating the rotational modulation of the light curve may be confined to a range of stellar latitudes that may not reflect the same velocity field encoded in the rotational broadening of spectral lines, and (2) the star may have an inclination resulting in the value of $\sin i_*$ being substantially <1 . If this is the case, the orbit of K2-33 b is misaligned with the spin of its host star. While $R_* \sin i_*$ provides a valuable consistency check on stellar radius, we use the Stefan–Boltzmann radius given its insensitivity to the unknown stellar inclination.

K2 time series photometry treatment. The K2 mission³⁷ observes fields along the ecliptic plane for approximately 75 days at a time. K2 photometry possesses percentage-level systematic signatures from pointing drift and intra-pixel detector sensitivity variations that must be corrected in order to detect sub-per-cent planet transits. We acquired such a corrected light curve from the Exoplanet Follow-up Observing Program public website⁹, derived using a $12'' \times 16''$ rectangular aperture (Extended Data Fig. 4).

Before modelling the transit profile, we removed the spot modulation pattern using a cubic basis spline fit with knots spaced by 12 long-cadence measurements,

employing iterative rejection of outliers³⁸. We verified that no in-transit observations were included in the spline fit by phasing the flattened light curve on the orbital period and inspecting the points excluded from the fit. An artificial gap in the systematics-corrected light curve from the ExoFOP page was not adequately captured by the spline fit and, consequently, we excluded from further analysis those data with BJD values in the range 2,456,936 to 2,456,938, resulting in the loss of a single transit. We assigned a constant observational uncertainty for each K2 measurement, determined from the standard deviation in the out-of-transit light curve (here defined as observations more than 12 h from either side of the transit centres).

Transit model-fitting analysis. We employed previously established methodology¹⁰ for fitting transit models to the light curve. The approach uses the BATMAN software³⁹ (<https://pypi.python.org/pypi/batman-package/2.1.0>), based on the Mandel and Agol analytic light-curve formalism⁴⁰, to generate model transit profiles. Transit models were numerically integrated to match the ~ 30 -min cadence of K2 observations. We assumed a linear limb-darkening law for the star, imposing a Gaussian prior on the linear limb-darkening coefficient, u , based on tabulated values⁴¹ appropriate for the effective temperature and surface gravity of K2-33. We also allowed for dilution by light from a second star in the fitting. In post-processing, we selected only those samples corresponding to dilution levels consistent with our companion exclusion analysis (Fig. 2).

The directly fitted transit parameters are the scaled semi-major axis (a/R_*), the planet-to-star radius ratio (R_p/R_*), the orbital period (P), the time of mid-transit (t_0) and the inclination (i). The multi-dimensional transit parameter space was explored using an affine invariant implementation of the Markov chain Monte Carlo algorithm⁴² to find the best-fit model and determine parameter uncertainties. Each observation was weighted equally, resulting in a best-fit likelihood of $-1/2 \times \chi^2$. We followed established methods¹⁰ for Markov chain Monte Carlo initialization, burn-in treatment, rescaling of data weights, and convergence testing. Table 1 quotes median transit model values, with uncertainties determined from the 15.87% and 84.13% percentiles of the parameter posterior distributions.

The scatter in transit is $\sim 15\%$ larger than the scatter in equal duration intervals before and after transit. One possible explanation may be spot-crossing events, given that the star is expected to have a high spot-covering fraction⁴³, supported by the modulation in the unflattened light curve.

Prior to adopting the publicly available light curve⁹, we independently extracted a light curve from K2 target pixel files using a different photometric pipeline^{10,44}. Performing the same modelling described above on this second light curve produced consistent results.

High-spectral-resolution observations and radial velocities. We used the High Resolution Spectrograph⁴⁵ (HIRES) on the 10-m Keck-1 telescope to measure the radial velocity of K2-33 relative to the Solar System barycentre (Extended Data Table 1) to confirm its cluster membership, and to constrain the mass of K2-33 b. The resolution of these spectra are $R \approx 50,000$ in the range $\sim 3,600\text{--}8,000$ Å for the 2016 epochs and $R \approx 36,000$ in the range $\sim 4,800\text{--}9,200$ Å for the 2015 epoch.

For the first epoch, velocity was derived by cross-correlating the spectrum with radial velocity standards⁴⁶ observed using HIRES in the same spectrograph configuration. Uncertainty is quantified from the dispersion among measurements relative to different standards, and over many different spectral orders. For the three epochs in 2016, systemic radial velocity was measured using the telluric A and B absorption bands as a fiducial wavelength reference. Assessing all measurements, the star's systemic radial velocity is -6.6 ± 0.1 km s⁻¹, where we quote a weighted average and standard deviation.

Limits on companions from the spectroscopic data. We searched for and excluded distant gravitationally bound companions to K2-33 by looking for trends in the 265-day radial velocity time series. A 3σ upper limit to any possible acceleration is 2.6 km s⁻¹ yr⁻¹. Following established methods⁴⁷, the limiting minimum mass detectable from the radial velocity measurements as a function of orbital separation rules out additional companions of $\geq 0.14M_\odot$ at 3 AU and $\geq 0.39M_\odot$ at 5 AU.

We also searched for secondary spectral lines⁴⁸ that would arise from a companion projected within $0.4''$ of the primary. No stars as faint as 3% of the brightness of the primary were detected, though we are blind to companion stars with a small (<15 km s⁻¹) radial velocity relative to the primary because the spectral lines of the two stars would not be distinguished.

High-resolution imaging. Using adaptive optics at the Keck-2 telescope on UTC 2011 May 15, we obtained ten 9-s exposures of K2-33. A second set of Keck/NIRC2 images was acquired on UTC 2016 February 17 in a three-point dither pattern with 10-s integrations per dither position, repeated three times. For both epochs, the narrow-camera optics were used resulting in a pixel scale of 9.942 mas per pixel; the final co-added images have a resolution of $0.07''$ (full-width at half-maximum). The observations at the two epochs have the same total integration time, so the final images have comparable depth.

To estimate sensitivity to point sources as a function of radial distance from the star, median flux levels and root-mean-square dispersions (σ) were calculated in incremental annuli centred on the source. An image was constructed with these characteristics and synthetic sources with full-width at half-maximum equivalent to that of K2-33 were injected into the image at varying positions and brightnesses. The synthetic sources were then measured to determine the 5σ detection limit. Comparing the instrumental magnitude to that of the star produces ΔK for that annulus.

Non-redundant aperture masking. K2-33 was observed using an aperture mask on the same 2011 night as the clear aperture images were taken. Aperture masking interferometry⁴⁹ uses an opaque mask with clear holes such that the baseline between any two samples a unique spatial frequency in the pupil plane, and achieves angular resolution as good as $1/3 \times \lambda/D$ compared to $1.2 \times \lambda/D$, though at the expense of diminished throughput^{50,51}.

With the nine-hole mask in the NIRC2 camera, we obtained 40 dithered images using 20-s exposures. Observations of calibrator stars were interspersed with the targets and obtained in an identical manner. In each image, the mask creates a set of 36 overlapping interference fringes on the detector. The bispectrum, the complex triple product of visibilities defined by the three baselines formed from any three subapertures, is then calculated. The phase of this complex quantity is the closure phase, which has the advantage of being largely insensitive to phase delays owing to atmospheric effects or residual uncorrected phase aberrations not sensed by the adaptive optics system. We followed established procedures^{52,53} for calculating the closure phase for calibrator stars.

Limits on companions from the imaging data. Using all three sets of high-spatial-resolution near-infrared imaging data, we searched for projected companions to K2-33 (either gravitationally bound or foreground/background sources). The two epochs of clear aperture data each achieved $\Delta K > 4.5$ mag of contrast beyond $0.13''$ and $\Delta K > 7.5$ mag beyond $1''$. In the interim between the observations, K2-33 moved on the sky by $0.1228'' \pm 0.0085''$ owing to its parallactic and proper motion. The combined set of infrared images thus rules out nearly all unassociated background stars bright enough to produce the transit observed in the optical. The aperture masking observations are sensitive to more closely projected (from $0.02''$ – $0.16''$) companions in the stellar and substellar mass regimes. We use these to rule out potential associated companions larger than 19 Jupiter masses down to orbital separations of 3 AU, and are sensitive to companions with masses as low as 11–12 Jupiter masses in the range 6–23 AU, where quoted companion masses are model-dependent¹⁹ conservatively assuming an age of 10 Myr. Finally, a prior analysis⁵⁴ of near-infrared photometry ruled out associated companions down to masses of 5%–6% of the primary mass at separations of $1.7''$ – $27.5''$, or approximately 250–4,000 AU.

The imaging limits constrain the brightnesses of any putative companions in the near-infrared K-band. To approximately convert these contrast limits to constraints in the optical Kepler bandpass, K_p , we employed a combination of theoretical evolutionary models and empirical colour–colour relations. The TRILEGAL simulation discussed below predicts that the mean potential contaminant towards K2-33 would have infrared colour $J - K_s = 0.57$ mag, corresponding to a K-type dwarf. From an empirical optical–infrared K_p -to- K_s conversion⁵⁵, we conclude that a typical contaminating source would have colour $K_p - K_s = 2.00$ mag. We thus gain an additional 2 magnitudes of contrast when converting the NIRC2 contrast curve to the corresponding limits in the Kepler bandpass when considering unassociated companions.

For associated companions, the imaging constraints natively derived in the near-infrared were converted to optical limits using pre-main-sequence evolutionary models¹⁹. Putative companion masses can be paired with our assumed primary mass to yield predicted R-band contrasts at 5–10 Myr, where the R-band serves as a proxy for the Kepler bandpass. Notably, for the clear aperture adaptive optics region of Fig. 2b, the contrast achieved beyond 30 AU is better than represented in the figure owing to limitations of the models, which do not extend below $10M_{\text{Jup}}$ (corresponding to $\Delta K_p \geq 9.5$ mag).

We rule out nearly all stellar and brown-dwarf mass companions to K2-33, with the exception of a narrow swath of parameter space corresponding to ultra-low-mass stars or brown dwarfs separated by 1–3 AU from the primary (Fig. 2). The physically permitted (as opposed to unexcluded) separation range of any hypothetical diluting companion is even smaller, considering that the inferred inner edge of the disk is at 2 AU.

Galactic structure model and intracluster contamination. In addition to searching directly for sources that could contaminate the K2 light curve, we estimated the expected surface density of such sources as a function of magnitude using the TRILEGAL version 1.6 model^{56,57} (<http://stev.oapd.inaf.it/cgi-bin/trilegal>) of the Milky Way Galaxy. Notably, TRILEGAL does not include the local extinction due to gas and dust associated with Upper Scorpius itself, and therefore produces an upper limit to the field-star source density. We simulated a 1-deg^2 field and scaled

the resulting numbers first to the $10 \times 10 \text{ arcsec}^2$ field of view of the Keck/NIRC2 images and the $12 \times 16 \text{ arcsec}^2$ K2 photometric aperture, and then to the unexcluded regions of parameter space in Fig. 2a. We found the TRILEGAL prediction for the expected surface density of sources to the $K_s < 18.0$ mag limit (5σ) of the NIRC2 data, and translated it to < 0.15 sources per NIRC2 field, consistent with our detection of none. Within the K2 photometry aperture, < 0.3 unassociated sources are expected to the same magnitude limit, corresponding to a mean optical brightness of $K_p = 18.6$ mag.

By considering the surface density of sources expected from integrating essentially all the way through the galaxy, a maximum of three sources are expected in the K2 aperture, having mean optical brightness $K_p = 23.1$ mag, which is too faint to explain the transit depth. Indeed, we have ruled out nearly all companions with projected separations larger than $0.04''$, as well as effectively all of those inside $0.04''$; fewer than 4×10^{-6} sources are expected in the remaining unexcluded parameter space of Fig. 2a. An even smaller number of sources are expected to be eclipsing stellar binaries. Therefore, in addition to not detecting any contaminants in the high-spatial-resolution imaging and spectroscopic data, we argue that essentially none are expected.

A similar source-density argument can be used to constrain the probability of contaminants with nearly identical proper motions to K2-33, probably association members that are foreground or background to K2-33. The multi-epoch adaptive optics imaging cannot rule out closely projected sources within $0.02''$ (the inner limit probed by the aperture masking observations) that are also co-moving with K2-33 between 2011 and 2016. From the observed Upper Scorpius mass function and width of the association on the sky¹², we estimated a source density of about 16 members deg^{-2} . Thus, fewer than 2×10^{-9} co-moving contaminants are expected within $0.02''$ of K2-33.

False-positive probability analysis. Eclipsing stellar binaries, when diluted by the light of a third star, can produce light curves that masquerade as a planetary transit. These false positives come in three broad classes: (1) undiluted eclipsing stellar binaries, (2) background (and foreground) eclipsing stellar binaries where the eclipses are diluted by the target star, and (3) bound eclipsing stellar binaries in hierarchical triple systems.

We used the VESPA program²³ (<https://pypi.python.org/pypi/VESPA/0.4.7>) to compare the likelihood of each binary scenario against the planetary interpretation. As input for the calculation, we provide the K2 light curve, the stellar parameters, and to be as conservative as possible, we adopt our least stringent imaging constraints: the 2011 NIRC2 clear contrast curve, and the aperture masking limits. Even in this minimally constraining scenario, we find a false-positive probability of $< 1 \times 10^{-11}$ from VESPA, as expected given our exclusion in Fig. 2 of essentially all the modelled scenarios. Notably, however, VESPA does not account for substellar objects, pre-main-sequence evolution, extinction, or the unknown prior probability of planets around 5–10-Myr-old stars.

Implications of hierarchical triple scenarios. The conjectured hierarchical triple configuration is argued elsewhere as unlikely on the basis of population statistics; however, we must still consider the possibility that K2-33 b has a larger radius owing to dilution of the transit depth by a luminous companion to K2-33. We first consider the case in which the planet orbits K2-33, but the transit depth is diluted by an undetected secondary. In this scenario, the ratio of the true planet radius to the observed planet radius (that is, not accounting for dilution) is $R_{p,\text{true}}/R_{p,\text{obs}} = \sqrt{1 + F_{\text{sec}}/F_{\text{pri}}}$, where $F_{\text{sec}}/F_{\text{pri}}$ is the optical flux ratio between the secondary and primary. For ΔK_p in the range 2–6 mag, the true planet radius is at most 7.6% larger, within our quoted uncertainties.

Now, we consider the case in which the planet orbits an undetected companion to K2-33. In this scenario, the planet radius is $R_p = R_{\text{sec}} \sqrt{\delta_0(1 + F_{\text{pri}}/F_{\text{sec}})}$, where R_{sec} is the secondary radius, and δ_0 is the observed transit depth. Using ΔR as a proxy for ΔK_p , we found the secondary radius implied by the optical brightness decrement using evolutionary models¹⁹ valid for 5–10 Myr. For ΔK_p in the range 2–6 mag, we found that the implied planet radius is in the range $(0.56\text{--}1.85)R_{\text{Jup}}$. At these ages, such radii correspond to planet masses of $\leq 13M_{\text{Jup}}$. Only for $\Delta K_p \geq 6$ mag does the implied mass exceed the nominal brown-dwarf minimum mass. However, coeval eclipsing brown dwarfs are unlikely to produce eclipse depths of $> 50\%$ (corresponding to contrasts of $\Delta K_p \geq 5.8$ mag). Furthermore, we argue that scenarios involving eclipsing brown dwarfs 1–3 AU from K2-33 are extremely unlikely, for several compounding reasons: (1) the observed frequency of brown-dwarf companions to M-dwarfs is a few per cent^{22,58}, (2) in the restricted domain of $a = 1\text{--}3$ AU, the frequency is lower still; and (3) the frequency of eclipsing brown dwarf pairs so similar in size and temperature that the primary and secondary eclipses are indistinguishable is smaller still. Finally, the tenuous dust disk with inner edge at 2 AU is further evidence in favour of the single-star scenario.

Cluster age. The age of Upper Scorpius is constrained to be 5–10 Myr, from a variety of considerations. Absence of dense molecular gas or protostars in Upper

Scorpius implies that star formation has ceased in the region⁵⁹, while the presence of protoplanetary disks around a significant fraction of members indicates that planet formation is ongoing¹³. However, the precise age of the association is currently debated. An early kinematic analysis, in which the motions of high-mass members were traced back in time to the point of closest proximity to one another, suggested an age of 5 Myr (ref. 60). The first Hertzsprung–Russell diagram analysis of the full stellar population, spanning from the highest to the lowest masses, also determined an age of 5 Myr without appreciable dispersion¹². Most subsequent age determinations using theoretical evolutionary models in the Hertzsprung–Russell diagram arrived at the same consistent age of 5–6 Myr for massive main-sequence turnoff stars^{61,62}, low-mass stars^{17,63–65}, as well as substellar mass objects⁶⁶.

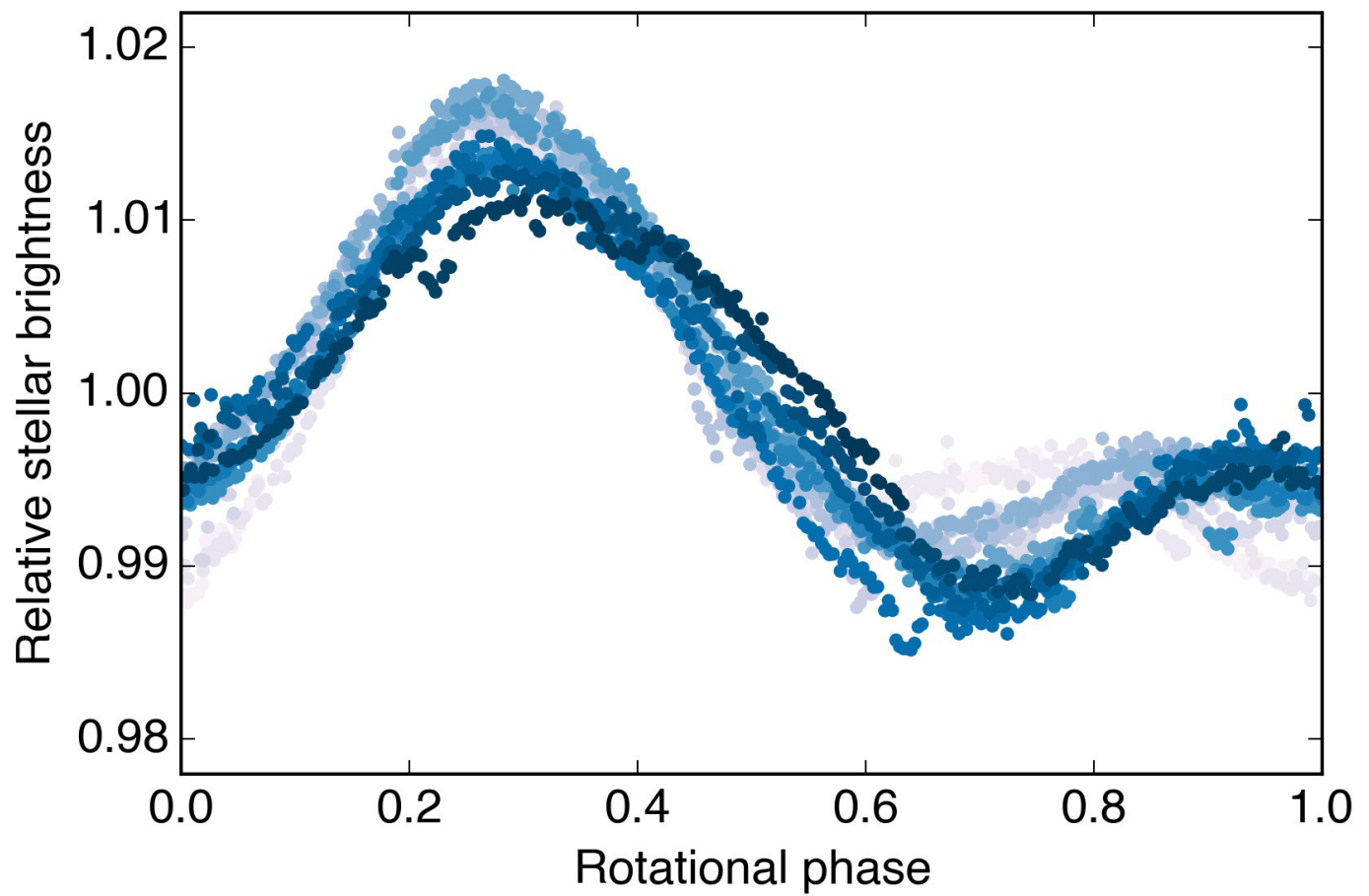
However, the association age has also been determined to be closer to 10 Myr from analyses of low-mass members^{67,68} and the intermediate-mass pre-main-sequence and main-sequence population, main-sequence turn-off stars, and the supergiant Antares⁶⁹. Emerging evidence from double-lined eclipsing binaries also supports an age in the range of 7–11 Myr^{38,70,71}, as does an updated kinematic analysis⁶⁹. Despite the lack of consensus on the precise age of Upper Scorpius, the full error-inclusive range of estimates in the literature (3–13 Myr) place the association at a critically important stage in the planet formation process—when most primordial disks have dispersed⁷².

K2-33 b in the context of other claimed young planets. While several secure short-period planets have been found in orbit around stars in benchmark open clusters including the 600–800-Myr-old Hyades^{73,74} and Praesepe^{75,76}, the evidence for planets at younger ages is mixed. Direct imaging has revealed (5–10) M_{Jup} ‘planetary mass companions’ located at large semi-major axes from several stars having ages of a few tens to a few hundreds of million years. Additionally, there are strong indications of ongoing planet formation in many 1–3-Myr-old circumstellar disks, based on the observed radial structure of dust.

However, there are no confirmed planets in the Jupiter or sub-Jupiter mass range with ages less than those corresponding to the late-heavy-bombardment era in our own Solar System. Both TW Hya and PTF0 8–8695 have been claimed to host hot Jupiter candidates detected via radial velocity or transit methods, but neither object has stood up to scrutiny^{77–79}. K2-33 b at an age of 5–10 Myr, in contrast, is a secure transiting planet. It is slightly larger than Neptune and its mass is probably similar to Neptune’s mass.

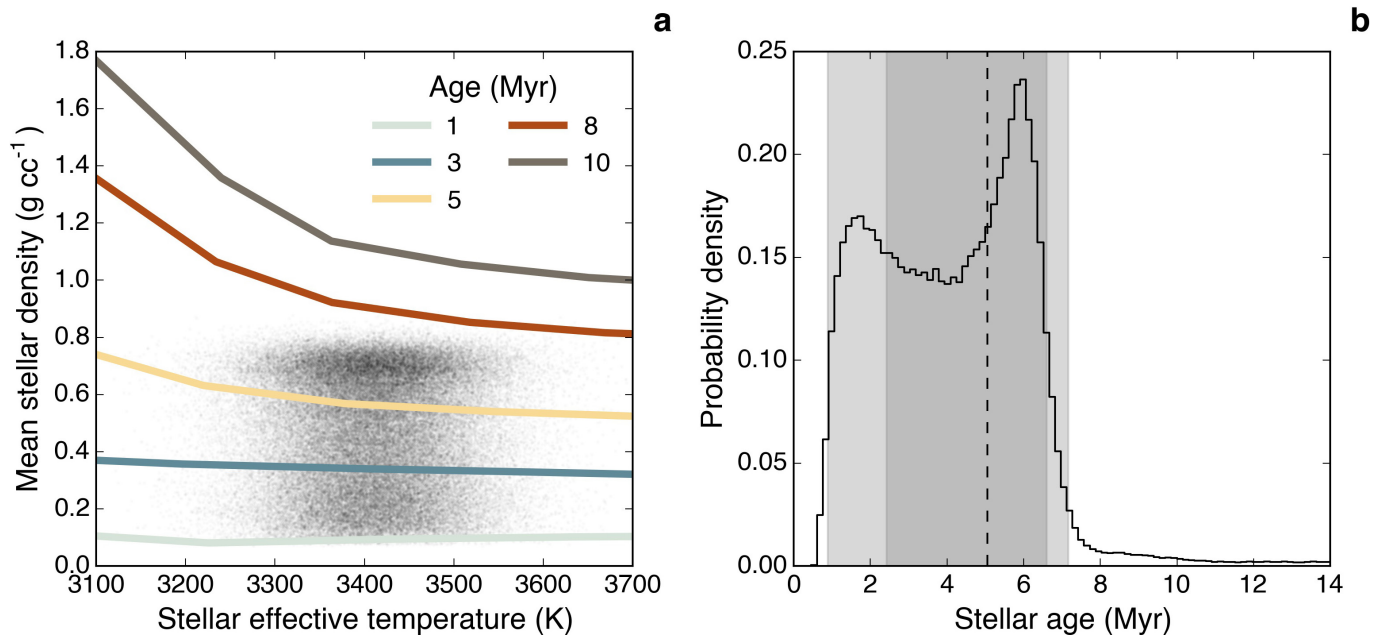
Code availability. We have opted not to make available our custom codes employed in the various elements of our data reduction and analysis, as they are either not of general purpose use, lack sufficient documentation, or they are already publicly available. The results of the computer processing are provided as Source Data. The raw data are publicly available, either currently (K2) or after a limited proprietary period (Keck, ALMA).

31. Lépine, S. & Simon, M. Nearby young stars selected by proper motion. I. Four new members of the β Pictoris moving group from the Tycho-2 catalog. *Astron. J.* **137**, 3632–3645 (2009).
32. Chen, C. H. et al. A Magellan MIKE and Spitzer MIPS study of 1.5–1.0 M_{\odot} stars in Scorpius–Centaurus. *Astrophys. J.* **738**, 122 (2011).
33. de Bruijne, J. H. J. Structure and colour-magnitude diagrams of Scorpius OB2 based on kinematic modelling of Hipparcos data. *Mon. Not. R. Astron. Soc.* **310**, 585–617 (1999).
34. Seager, S. & Mallén-Ornelas, G. A unique solution of planet and star parameters from an extrasolar planet transit light curve. *Astrophys. J.* **585**, 1038–1055 (2003).
35. Lomb, N. R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39**, 447–462 (1976).
36. Scargle, J. D. Studies in astronomical time series analysis. II—Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982).
37. Howell, S. B. et al. The K2 mission: characterization and early results. *Publ. Astron. Soc. Pacif.* **126**, 398–408 (2014).
38. David, T. J., Hillenbrand, L. A., Cody, A. M., Carpenter, J. M. & Howard, A. W. K2 discovery of young eclipsing binaries in Upper Scorpius: direct mass and radius determinations for the lowest mass stars and initial characterization of an eclipsing brown dwarf binary. *Astrophys. J.* **816**, 21 (2015).
39. Kreidberg, L. BATMAN: BASIC Transit Model cALculationN in Python. *Publ. Astron. Soc. Pacif.* **127**, 1161–1165 (2015).
40. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
41. Claret, A., Hauschildt, P. H. & Witte, S. New limb-darkening coefficients for PHOENIX/1D model atmospheres. I. Calculations for 1500 K $\leq T_{\text{eff}} \leq 4800$ K Kepler, CoRoT, Spitzer, uvby, UBVRIJK, Sloan, and 2MASS photometric systems. *Astron. Astrophys.* **546**, A14 (2012).
42. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
43. Jackson, R. J. & Jeffries, R. D. The effect of starspots on the radii of low-mass pre-main-sequence stars. *Mon. Not. R. Astron. Soc.* **441**, 2111–2123 (2014).
44. Petigura, E. A. et al. Two transiting Earth-size planets near resonance orbiting a nearby cool star. *Astrophys. J.* **811**, 102 (2015).
45. Vogt, S. S. et al. In *Instrumentation in Astronomy VIII* (eds Crawford, D. L. & Craine, E. R.) *Proceedings of SPIE* Vol. 2198, 362–375, <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=959834> (SPIE, 1994).
46. Nidever, D. L., Marcy, G. W., Butler, R. P., Fischer, D. A. & Vogt, S. S. Radial velocities for 889 late-type stars. *Astrophys. J. Suppl. Ser.* **141**, 503–522 (2002).
47. Montet, B. T., Crepp, J. R., Johnson, J. A., Howard, A. W. & Marcy, G. W. The TRENDS high-contrast imaging survey. IV. The occurrence rate of giant planets around M dwarfs. *Astrophys. J.* **781**, 28 (2014).
48. Kolbi, R., Marcy, G. W., Isaacson, H. & Howard, A. W. Detection of stars within $\sim 0.8''$ of Kepler objects of interest. *Astron. J.* **149**, 18 (2014).
49. Tuthill, P. G., Monnier, J. D., Danchi, W. C., Wishnow, E. H. & Haniff, C. A. Michelson interferometry with the Keck I telescope. *Publ. Astron. Soc. Pacif.* **112**, 555–565 (2000).
50. Baldwin, J. E., Haniff, C. A., Mackay, C. D. & Warner, P. J. Closure phase in high-resolution optical imaging. *Nature* **320**, 595–597 (1986).
51. Readhead, A. C. S. et al. Diffraction-limited imaging with ground-based optical telescopes. *Astron. J.* **95**, 1278–1296 (1988).
52. Kraus, A. L., Ireland, M. J., Martinache, F. & Hillenbrand, L. A. Mapping the shores of the brown dwarf desert. II. Multiple star formation in Taurus–Auriga. *Astrophys. J.* **731**, 8 (2011).
53. Hinkley, S. et al. Discovery of seven companions to intermediate-mass stars with extreme mass ratios in the Scorpius–Centaurus association. *Astrophys. J.* **806**, L9 (2015).
54. Kraus, A. L. & Hillenbrand, L. A. Multiple star formation to the bottom of the initial mass function. *Astrophys. J.* **757**, 141 (2012).
55. Howell, S. B. et al. Kepler-21b: a 1.6 R_{Earth} planet transiting the bright oscillating F subgiant star HD 179070. *Astrophys. J.* **746**, 123 (2012).
56. Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E. & da Costa, L. Star counts in the Galaxy. Simulating from very deep to very shallow photometric surveys with the TRILEGAL code. *Astron. Astrophys.* **436**, 895–915 (2005).
57. Girardi, L. et al. TRILEGAL, a TRIdimensional model of the GALaxy: status and future. *Astrophys. Space Sci. Proc.* **26**, 165–170 (2012).
58. Dieterich, S. B., Henry, T. J., Golimowski, D. A., Krist, J. E. & Tanner, A. M. The solar neighborhood. XXVIII. The multiplicity fraction of nearby stars from 5 to 70 AU and the brown dwarf desert around M dwarfs. *Astron. J.* **144**, 64 (2012).
59. Preibisch, T. & Mamajek, E. *The Nearest OB Association: Scorpius–Centaurus (Sco OB2)* (ed. Reipurth, B.) 235–284, <http://aspmonographs.org/custom/publications/paper/005-0235.html> (Astronomical Society of the Pacific Monograph Publications, 2008).
60. Blaauw, A. in *NATO Advanced Science Institutes Series C* (eds Lada, C. J. & Kylafis, N. D.) Vol. 342, 125–154 (ASI, 1991).
61. de Zeeuw, T. & Brand, J. in *Birth and Evolution of Massive Stars and Stellar Groups* (eds Boland, W. & van Woerden, H.) *Astrophysics and Space Science Library* Vol. 120, 95–101 (Springer, 1985).
62. de Geus, E. J., de Zeeuw, P. T. & Lub, J. Physical parameters of stars in the Scorpio–Centaurus OB association. *Astron. Astrophys.* **216**, 44–61 (1989).
63. Preibisch, T. & Zinnecker, H. The history of low-mass star formation in the Upper Scorpius OB association. *Astron. J.* **117**, 2381–2397 (1999).
64. Slesnick, C. L., Hillenbrand, L. A. & Carpenter, J. M. A large-area search for low-mass objects in Upper Scorpius. II. Age and mass distributions. *Astrophys. J.* **688**, 377–397 (2008).
65. Lodieu, N., Dobbie, P. D. & Hambly, N. C. Multi-fibre optical spectroscopy of low-mass stars and brown dwarfs in Upper Scorpius. *Astron. Astrophys.* **527**, A24 (2011).
66. Allen, P. R., Trilling, D. E., Koerner, D. W. & Reid, I. N. Luminosity functions of young clusters: modeling the substellar mass regime. *Astrophys. J.* **595**, 1222–1230 (2003).
67. Martin, E. L. Weak and post-T Tauri stars around B-type members of the Scorpius–Centaurus OB association. *Astron. J.* **115**, 351–357 (1998).
68. Frink, S. *Kinematics of T Tauri Stars in Nearby Star Forming Regions*. PhD thesis, Astronomisches Rechen-Institut Heidelberg (1999); <http://adsabs.harvard.edu/abs/1999PhDT.....9F>.
69. Peca, M. J., Mamajek, E. E. & Bubar, E. J. A revised age for Upper Scorpius and the star formation history among the F-type members of the Scorpius–Centaurus OB association. *Astrophys. J.* **746**, 154 (2012).
70. Kraus, A. L. et al. The mass-radius relation of young stars. I. USco 5, an M4.5 eclipsing binary in Upper Scorpius observed by K2. *Astrophys. J.* **807**, 3 (2015).
71. Lodieu, N. et al. An eclipsing double-line spectroscopic binary at the stellar/substellar boundary in the Upper Scorpius OB association. *Astron. Astrophys.* **584**, A128 (2015).
72. Wyatt, M. C. Evolution of debris disks. *Annu. Rev. Astron. Astrophys.* **46**, 339–383 (2008).
73. Mann, A. W. et al. Zodiacal exoplanets in time (ZEIT). I. A Neptune-sized planet orbiting an M4.5 dwarf in the Hyades Star Cluster. *Astrophys. J.* **818**, 46 (2016).
74. David, T. J. et al. New Pleiades eclipsing binaries and a Hyades transiting system identified by K2. *Astron. J.* **151**, 112 (2016).
75. Quinn, S. N. et al. Two “b”s in the beehive: the discovery of the first hot Jupiters in an open cluster. *Astrophys. J.* **756**, L33 (2012).
76. Malavolta, L. et al. The GAPS programme with HARPS-N at TNG. XI. Pr 0211 in M 44: the first multi-planet system in an open cluster. *Astron. Astrophys.* **588**, A118 (2016).
77. Rucinski, S. M. et al. Photometric variability of the T Tauri star TW Hya on time-scales of hours to years. *Mon. Not. R. Astron. Soc.* **391**, 1913–1924 (2008).
78. Huélamo, N. et al. TW Hydrae: evidence of stellar spots instead of a hot Jupiter. *Astron. Astrophys.* **489**, L9–L13 (2008).
79. Yu, L. et al. Tests of the planetary hypothesis for PTF0 8–8695b. *Astrophys. J.* **812**, 48 (2015).



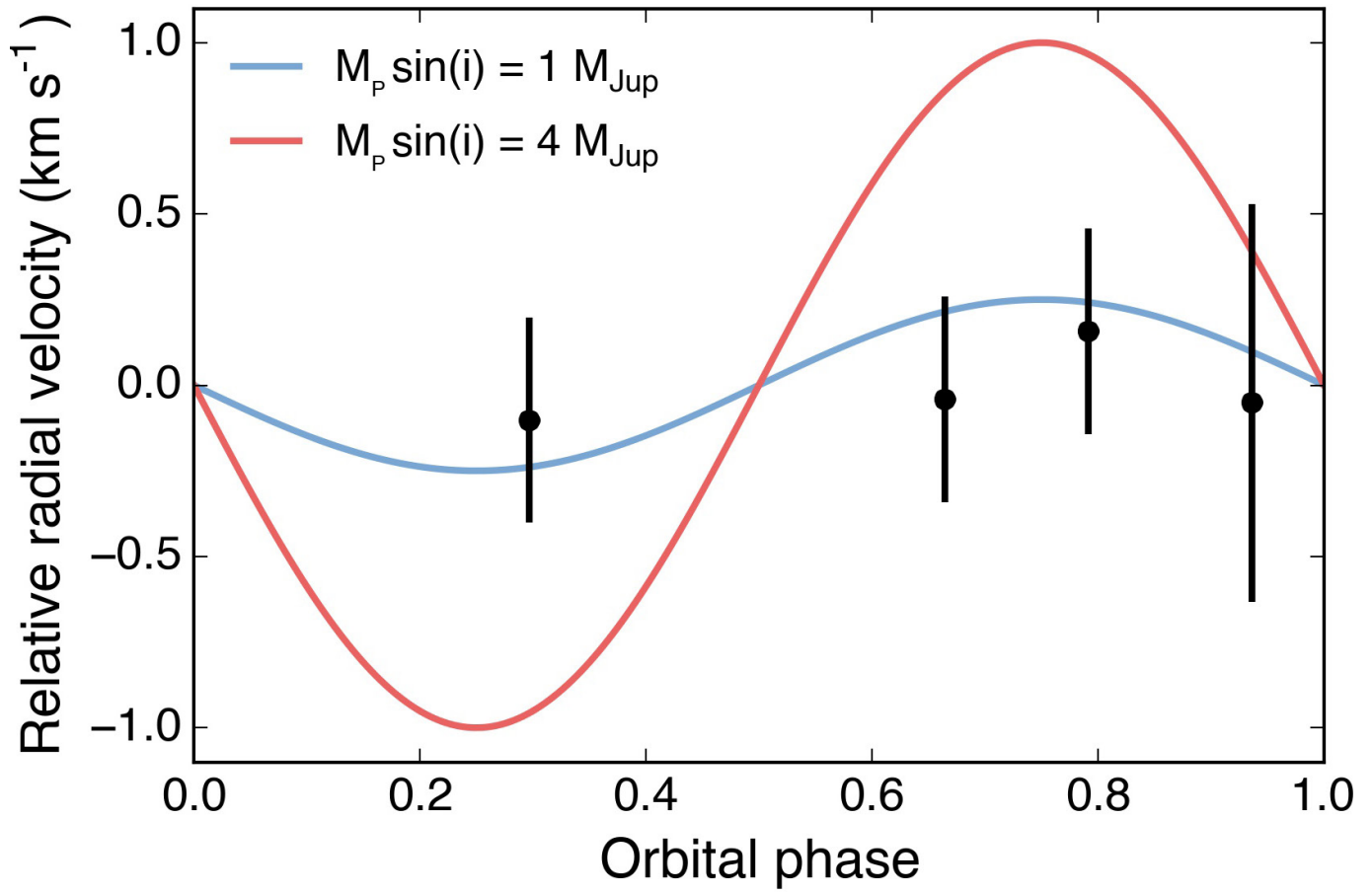
Extended Data Figure 1 | K2 light curve for K2-33 phased on the stellar rotation period of 6.3 days. Semi-sinusoidal brightness variations due to rotational modulation of starspots. Point colour indicates the relative time of observation, with grey corresponding to earlier in the campaign and dark blue corresponding to later times. Brightness is lowest when the most heavily spotted hemisphere of the stellar surface is along the line of sight.

The shape and evolution of the variability pattern depends on the number, geometry, distribution, and lifetime of spots, along with any latitudinal gradient in the rotational speed (differential rotation). The transits of K2-33 b are visible by eye in this figure and are too narrow in rotational phase to be attributed to any feature on or near the stellar surface.



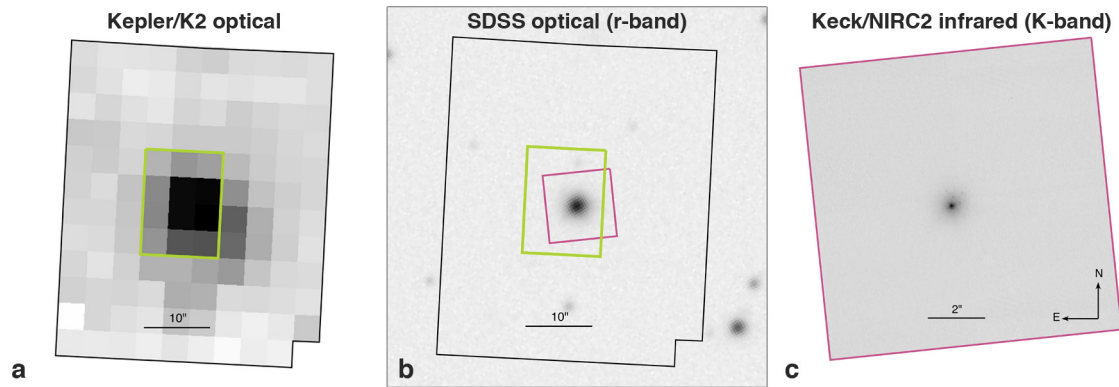
Extended Data Figure 2 | Model-dependent age of K2-33. **a**, Solid lines show mean stellar density as a function of effective temperature for pre-main-sequence stars having different ages, according to theoretical models¹⁹. Grey points represent plausible combinations of density and temperature for K2-33 as determined by light-curve fits and stellar

spectroscopy. **b**, Distribution of implied stellar age based on temperature, density, and pre-main-sequence models. The implied age of 2–7 Myr is consistent with the age we adopted of 5–10 Myr, derived independently. Dark- and light-grey shaded regions indicate 68% and 95% confidence intervals, respectively.



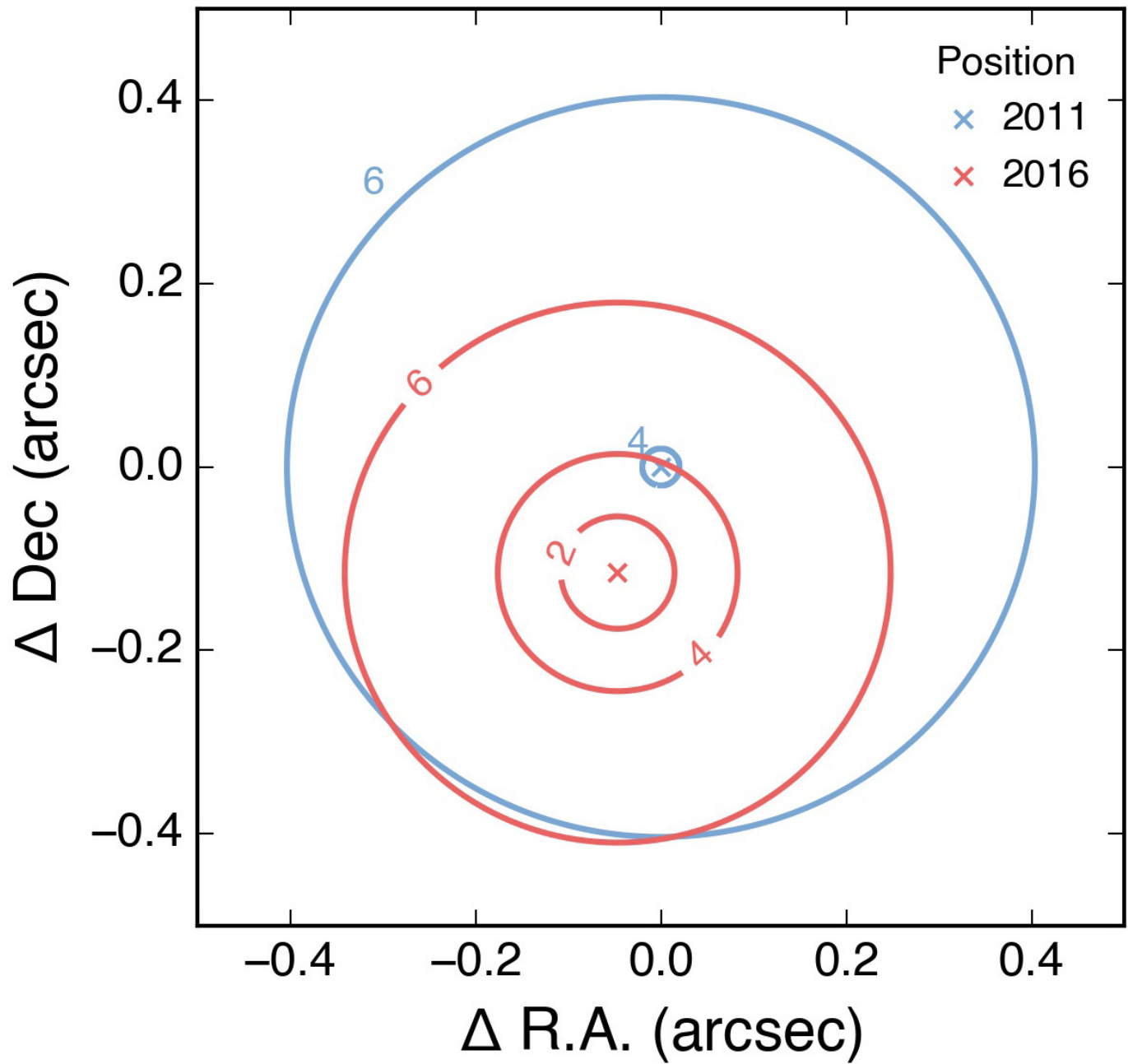
Extended Data Figure 3 | Apparent radial velocity variations of K2-33. Line-of-sight velocities and 1σ uncertainties (standard deviations, indicated by error bars) with respect to the Solar System barycentre from Keck/HIRES are indicated. Radial velocities are mean-subtracted, and the abscissa shows the orbital phase of K2-33 b measured from K2 photometry (mid-transit occurs at zero orbital phase). We rule out radial velocity

variations larger than 300 m s^{-1} at 68.3% confidence, corresponding to a $1.2 M_{\text{Jup}}$ planet mass. Curves show the expected radial velocity variations for planets having circular orbits and different masses M_p . Radial velocities due to a $1.0 M_{\text{Jup}}$ planet (blue) are consistent with our observations, while a $4.0 M_{\text{Jup}}$ planet (red) is ruled out at high confidence.



Extended Data Figure 4 | Images of K2-33. **a**, K2 target pixel file. **b**, Sloan Digital Sky Survey (SDSS) optical image. **c**, Keck/NIRC2 K-band image. Extents of the K2 target pixel file, K2 photometric aperture, and NIRC2 image are shown respectively with black, green, and purple boundaries. In each image, north is up and east is left. Three other sources identified by

SDSS reside within the K2 photometric aperture, one of which is a galaxy. All are 7.3–10.1 magnitudes fainter than K2-33 in the SDSS r-filter and below the detection limit of the NIRC2 images, and are thus too faint to produce the observed transits.



Extended Data Figure 5 | Sensitivity to non-comoving sources in the vicinity of K2-33. The blue X marks the star's position in 2011. Between 2011 and 2016, the star moved by $0.1228'' \pm 0.0085''$ (red X) owing to proper motion. Contours show the K-band sensitivity to non-comoving stars from adaptive optics imaging from both epochs. The 2011 data

set included non-redundant aperture masking, and provided tighter constraints. The combined sensitivity to non-comoving objects is the maximum contrast achieved for either data set. Owing to stellar proper motion, we achieved K-band contrasts of >3.3 mag throughout the $\Delta\text{RA}-\Delta\text{dec.}$ plane, even at the 2011 and 2016 positions of K2-33.

Extended Data Table 1 | Keck/HIRES radial velocities for K2-33

Epoch (UTC)	BJD	Radial velocity (km s ⁻¹)
2015-06-01 12:57:36.0	2457175.040	-6.61 ± 0.58
2016-02-02 15:30:14.4	2457421.146	-6.66 ± 0.30
2016-02-04 15:23:02.4	2457423.141	-6.60 ± 0.30
2016-02-21 14:29:45.6	2457440.104	-6.40 ± 0.30

Line-of-sight velocities and 1σ uncertainties (standard deviations) with respect to the Solar System barycentre. UTC, Coordinated Universal time.

A hot Jupiter orbiting a 2-million-year-old solar-mass T Tauri star

J. F. Donati^{1,2}, C. Moutou³, L. Malo³, C. Baruteau^{1,2}, L. Yu^{1,2}, E. Hébrard⁴, G. Hussain⁵, S. Alencar⁶, F. Ménard^{7,8}, J. Bouvier^{7,8}, P. Petit^{1,2}, M. Takami⁹, R. Doyon¹⁰ & A. Collier Cameron¹¹

Hot Jupiters are giant Jupiter-like exoplanets that orbit their host stars 100 times more closely than Jupiter orbits the Sun. These planets presumably form in the outer part of the primordial disk from which both the central star and surrounding planets are born, then migrate inwards and yet avoid falling into their host star¹. It is, however, unclear whether this occurs early in the lives of hot Jupiters, when they are still embedded within protoplanetary disks², or later, once multiple planets are formed and interact³. Although numerous hot Jupiters have been detected around mature Sun-like stars, their existence has not yet been firmly demonstrated for young stars^{4–6}, whose magnetic activity is so intense that it overshadows the radial velocity signal that close-in giant planets can induce. Here we report that the radial velocities of the young star V830 Tau exhibit a sine wave of period 4.93 days and semi-amplitude 75 metres per second, detected with a false-alarm probability of less than 0.03 per cent, after filtering out the magnetic activity plaguing the spectra. We find that this signal is unrelated to the 2.741-day rotation period

of V830 Tau and we attribute it to the presence of a planet of mass 0.77 times that of Jupiter, orbiting at a distance of 0.057 astronomical units from the host star. Our result demonstrates that hot Jupiters can migrate inwards in less than two million years, probably as a result of planet–disk interactions².

Very few exoplanets have yet been discovered around young, forming Sun-like stars aged less than ten million years (Myr)^{7,8}—called the T Tauri stars—either through the radial velocity variations or the photometric transits they induce in the light of their stars. Yet detections of young planets are key for our understanding of how planetary systems form and this is especially true of young hot Jupiters, which are thought to have a critical impact on the early architecture of these systems. The first claimed detection of a young hot Jupiter⁴ orbiting a T Tauri star was quickly refuted; the reported periodic radial velocity fluctuations were finally attributed to activity⁵ and to cool spots at the stellar surface⁹. The recent candidate detection of a transiting hot Jupiter around a T Tauri star⁶ is still pending confirmation.

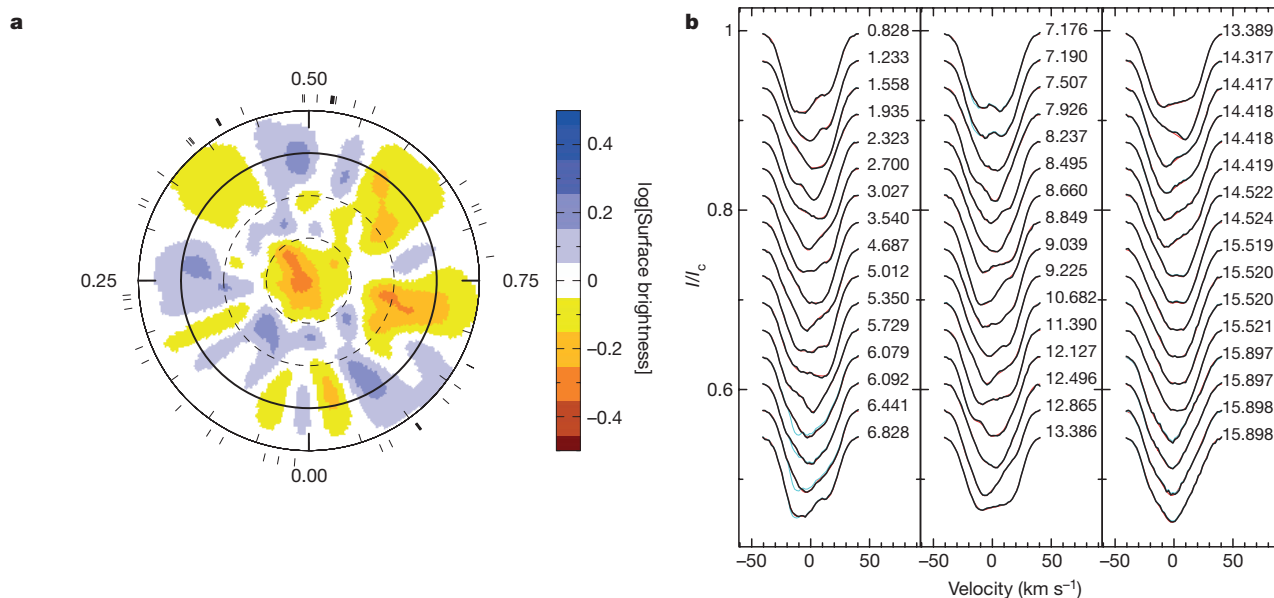


Figure 1 | Brightness map of V830 Tau and fit to the LSD profiles. **a**, Logarithmic brightness at the surface of V830 Tau as derived with Doppler imaging. Cool spots show as brown features and bright plages show as blue features. The rotation axis of the star is tilted at 55° to the line of sight, and the projected equatorial rotation velocity is equal to 30.5 km s^{−1} (ref. 12). The star is shown in a flattened polar view, with the pole in the centre and the equator depicted as a bold circle. Ticks

outside the image mark the phases of observations. **b**, Observed (black line) and modelled (red) LSD profiles of V830 Tau throughout our November–December 2015 run. LSD profiles before their filtering from lunar contamination are also shown (cyan lines). Numbers on the right of each profile indicate the rotation cycle. Cycles 6.079 to 7.190 are the most affected, contamination being much smaller or negligible in all other observations. I/I_c , normalized intensity.

¹Université de Toulouse, UPS-OMP, IRAP, 14 avenue Belin, F-31400 Toulouse, France. ²Centre National de la Recherche Scientifique (CNRS), IRAP/UMR 5277, 14 avenue Belin, F-31400 Toulouse, France. ³Canada–France–Hawaii Telescope (CFHT) Corporation, 65-1238 Mamalahoa Highway, Kamuela, Hawaii 96743, USA. ⁴Department of Physics and Astronomy, York University, Toronto, Ontario L3T 3R1, Canada. ⁵European Southern Observatory (ESO), Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany. ⁶Departamento de Física, ICEx, UFMG, avenida Antonio Carlos, 6627, 30270-901 Belo Horizonte, MG, Brazil. ⁷Université Grenoble Alpes, IPAG, BP 53, F-38041 Grenoble Cedex 09, France. ⁸CNRS, IPAG/UMR 5274, BP 53, F-38041 Grenoble Cedex 09, France. ⁹Institute of Astronomy and Astrophysics, Academia Sinica, PO Box 23-141, 106 Taipei, Taiwan. ¹⁰Département de physique, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada. ¹¹SUPA, School of Physics and Astronomy, University of St Andrews, St Andrews, KY16 9SS, UK.

T Tauri stars are known to harbour cool spots and bright features (plages) on their surfaces, generating radial velocity fluctuations with semi-amplitudes of several kilometres per second¹⁰, that is, much larger than the perturbations expected from a putative planet, even for close-in massive hot Jupiters inducing typical radial velocity signals of $\sim 0.1 \text{ km s}^{-1}$. Detecting hot Jupiters around T Tauri stars through velocimetry or photometry is thus quite challenging and requires efficient tools for filtering out the dominant jitter that activity induces in the spectra and light curves of young stars. We recently proposed a new method to achieve this goal¹¹, whose first applications to non-accreting (weak-line) T Tauri stars proved promising though inconclusive¹².

V830 Tau is a ~ 2 -Myr-old solar-mass T Tauri star¹² contracting towards the main sequence and currently spinning once in 2.741 days¹³, that is, ~ 10 times faster than the Sun. Evolutionary models¹⁴ suggest that it is fully or largely convective. Unlike 80% of the T Tauri stars in the Taurus star-forming region¹⁵, V830 Tau exhibits no significant infrared excess, implying that most of its inner accretion disk has already dissipated. This is consistent with its status as a non-accreting

weak-line T Tauri star, and makes it an ideal place to look for the presence of hot Jupiters at an early stage of star and planet formation.

In late 2015, we collected 48 high-resolution spectra of V830 Tau (see Extended Data Table 1) as part of the MaTYSSSE Large Program aimed at detecting hot Jupiters around weak-line T Tauri stars¹¹. Applying least-squares deconvolution¹⁶ (LSD) to our spectra, we derived accurate average line profiles and their temporal modulation over ~ 15 rotation cycles. Longitudinal magnetic fields were also derived from our circularly polarized data and the Zeeman signatures that fields generate in spectral lines¹⁶. Using tomographic techniques inspired from medical imaging, one can reconstruct distributions of spots and plages at the surfaces of rotating cool active stars from sets of densely sampled line profiles covering several rotation cycles. This method, called Doppler imaging¹⁷, can also probe the photospheric shear associated with surface differential rotation through the amount of twisting it generates in brightness maps^{18,19}.

Our Doppler imaging code was previously applied to a small set of 15 LSD profiles of V830 Tau, from which the distribution of surface features and the differential rotation pattern were recovered¹²; it even

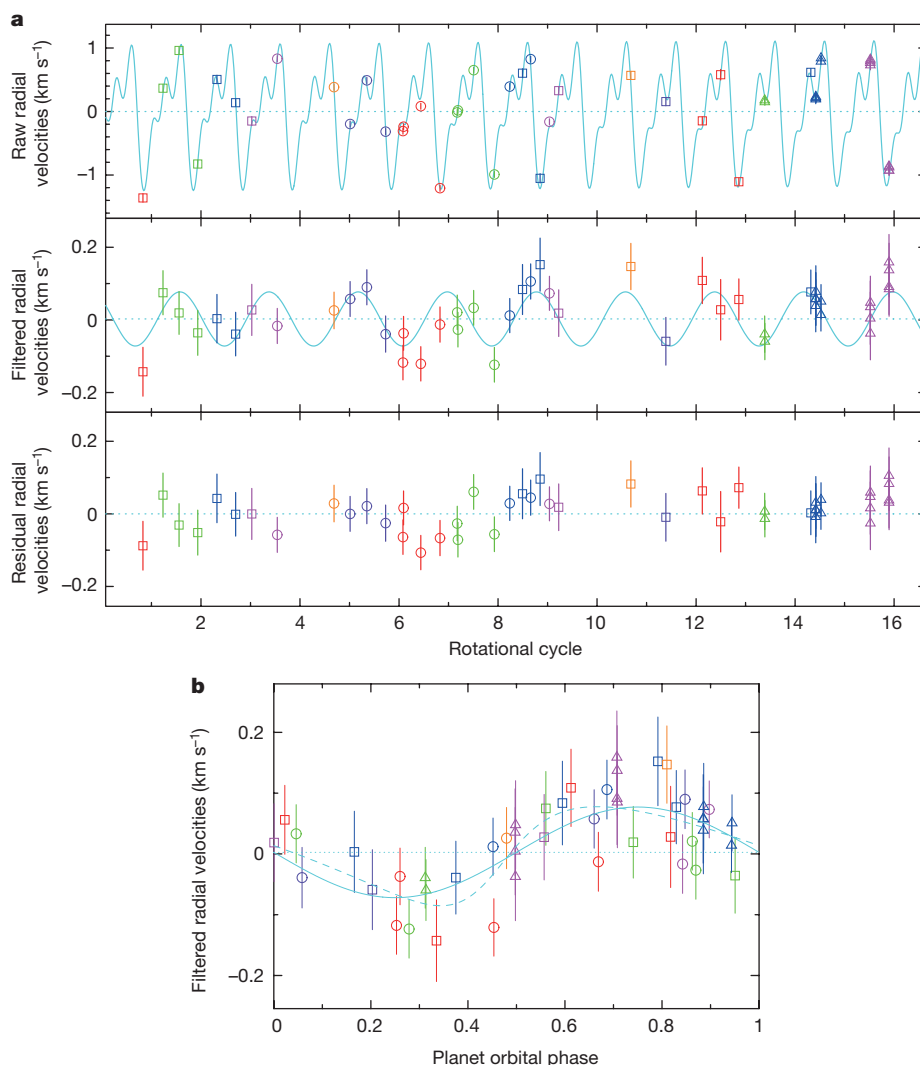


Figure 2 | Raw, filtered and residual radial velocities of V830 Tau.

a, Top, raw radial velocities of V830 Tau (open symbols and 1σ error bars) and the model inferred with Doppler imaging (cyan line); the model slowly evolves with time as a result of differential rotation. Open circles, squares and triangles depict ESPaDOs, NARVAL and ESPaDOs/GRACES data. Middle, activity-filtered radial velocities and best sine fit to the data (cyan line). The period and semi-amplitude of the planet radial velocity signal are equal to 4.93 ± 0.05 days and $75 \pm 11 \text{ m s}^{-1}$ (1σ error bars).

Bottom, residual radial velocities once the planet signal is removed and activity is filtered out, with a final root-mean-square dispersion of 48 m s^{-1} . Different colours code different rotation cycles. **b**, Activity-filtered radial velocities (with 1σ error bars) phase-folded on the planet orbital period of 4.93 days. Although the fit to the data is marginally better with an eccentric orbit (dashed line) than with a circular orbit (solid line), the significance of the derived eccentricity (0.30 ± 0.15) is too low to be reliable²³.

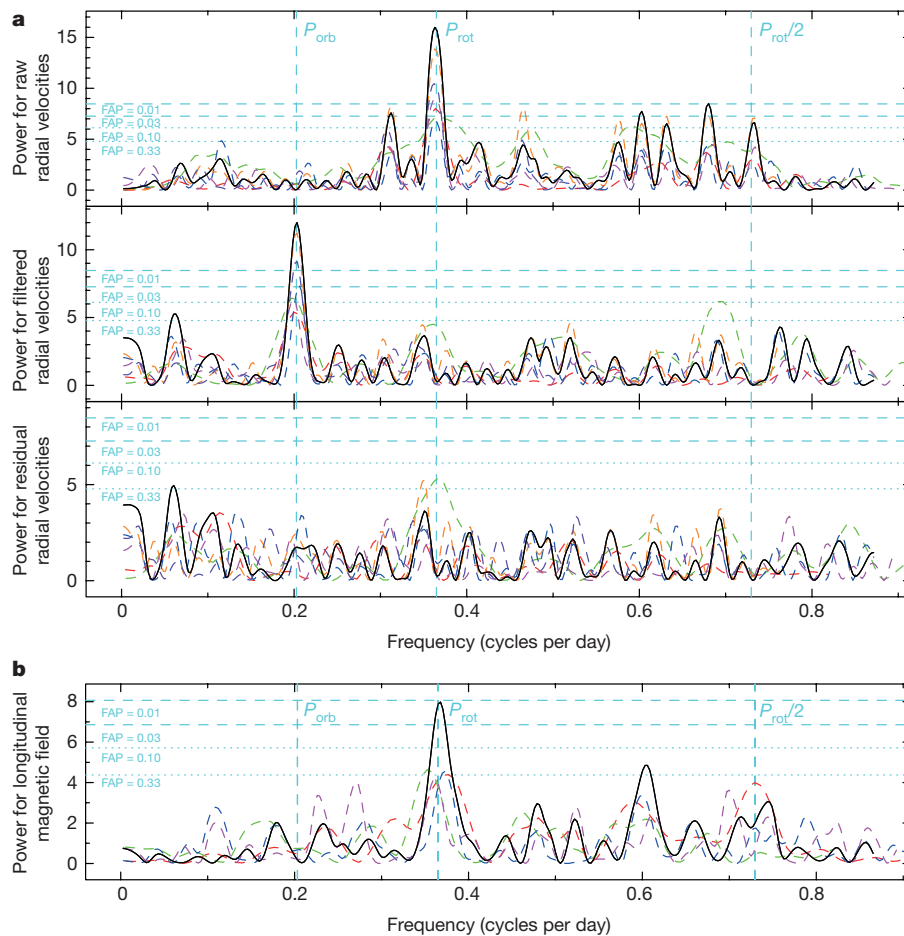


Figure 3 | Periodogram of the radial velocities and activity of V830 Tau.

a, Periodograms of the raw (top), filtered (middle) and residual (bottom) radial velocities shown in Fig. 2a. The black line is for the full set, while the dashed red, green, blue and pink lines are for the first half, the second half, the even points and the odd points only; the purple dashed lines are the periodograms once the 15 radial velocity points affected by lunar contamination are removed and the orange dashed lines are the periodograms once the 6 radial velocity points strongly affected by lunar contamination are removed. (Similar results are obtained when analysing subsets of filtered radial velocities from a global Doppler imaging

modelling, or when Doppler imaging filtering is applied to subsets of LSD profiles). The stellar rotation period (2.741 days), its first harmonic and the planet orbital period (4.93 days) are depicted with vertical dashed lines. The horizontal dotted and dashed lines trace the 33%, 10%, 3% and 1% false alarm probabilities (FAP). The planet signal in the filtered radial velocities is detected in the full set with a FAP < 0.03%. **b**, Periodogram of the line-of-sight projected (longitudinal) magnetic field, a reliable activity proxy²², featuring a clear peak at the stellar rotation period but no power at the planet orbital period.

suggested the potential presence of a hot Jupiter, though with a very low confidence level. The version of the Doppler imaging code used here implements a novel technique to filter out lunar contamination (which plagues spectra collected in non-photometric conditions), yielding good results when phase coverage is dense. Applying the code to our new set of 48 LSD profiles of V830 Tau yields the map and fit shown in Fig. 1. We again clearly detect differential rotation (see Extended Data Fig. 1) and confirm that it is ~ 3 times weaker than that of the Sun, reflecting that V830 Tau is largely or fully convective²⁰.

From the brightness image reconstructed with Doppler imaging, we derive the model radial velocity curve that V830 Tau should exhibit if all profile perturbations were attributable to surface features and differential rotation. By subtracting these modelled radial velocities from the observed ones (both computed as the first moment of the LSD profiles), we obtain the activity-filtered radial velocities of V830 Tau, whose amplitude is typically ~ 10 times lower than the raw radial velocities (see Fig. 2). A clear radial velocity signal, with a period of 4.93 ± 0.05 days and a semi-amplitude of $75 \pm 11 \text{ m s}^{-1}$, is detected in the activity-filtered radial velocities at a confidence level > 99.97% (see Figs 2 and 3). Using a Bayesian approach²¹, we find that the 4.93-day peak is at least 10^5 times more likely than any of the other features in the periodogram (see Extended Data Fig. 2a). The regular phase coverage

of our data also allows us to check that the 4.93-day signal is present in smaller subsets (for example, first and second half, even and odd points) though, as expected, with a lower confidence level. Similarly, we checked that our detection holds when profiles affected by lunar contamination are excluded (see Fig. 3 and Extended Data Fig. 2a) and when differential rotation is neglected. Periodograms of the longitudinal fields and of the H α emission, both reliable proxies for the activity jitter plaguing radial velocity curves²², show no power at a period of 4.93 days (see Fig. 3b and Extended Data Fig. 2b), demonstrating that the signal we report is unrelated to activity.

We interpret this radial velocity signal as being caused by a giant planet of mass 0.77 ± 0.15 Jupiter masses in a circular orbit around V830 Tau at a distance of 0.057 ± 0.001 AU (see Extended Data Table 2). Although the filtered radial velocities are marginally better fitted for an eccentric orbit ($e = 0.30 \pm 0.15$, see Fig. 2), we still favour a circular orbit given the large error bar on the eccentricity²³. Removing the planet signal from the original data and repeating the activity-filtering process yields residual radial velocities with a root-mean-square dispersion of 48 m s^{-1} (that is, consistent with the average noise level, see Extended Data Table 1) and no peak with a FAP < 30% left in the periodogram. Using the alternative option of fitting both the brightness map and the planet parameters simultaneously²⁴ yields identical results

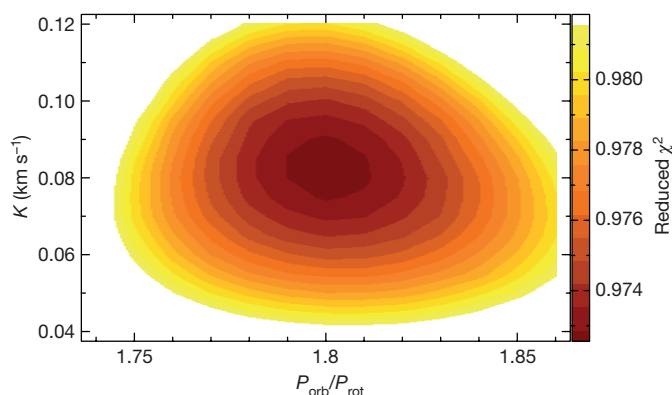


Figure 4 | Adjusting the planet parameters while modelling activity. Variations in the reduced χ^2 (shown as the colour scale) of the Doppler imaging fit to the LSD profiles for a given level of spottedness and assuming the presence of a planet in circular orbit, for a range of orbital periods P_{orb} and semi-amplitudes K of the radial velocity signature. (This is a two-dimensional cut from a three-dimensional map, with the phase of the radial velocity signal also included as a search parameter). The location of the minimum and local paraboloid curvature yield the optimal values of P_{orb} and K and their respective 1σ error bars²⁴, equal to 4.94 ± 0.05 days and 82 ± 10 m s⁻¹, in agreement with the results of our main filtering technique. The outer colour contour traces the projected 99.99% confidence interval, corresponding to a χ^2 increase of 21.1 for a three-parameter fit to the 2,208 data points of the LSD profiles. With respect to our best model incorporating a planet, a model with no planet corresponds to a χ^2 increase of 82, implying that the planet is detected with a FAP $< 10^{-15}$, much smaller than the FAP derived from the periodogram of the 48 radial velocity points (see Fig. 3) thanks to the larger number of data points in the fitted LSD profiles.

and demonstrates that our optimal model including a planet is orders of magnitude more likely than a model with no planet (see Fig. 4). Simulations in conditions identical to those of our observations yield results in close agreement with those of Figs 2 and 3 (see Extended Data Figs 3 and 4), further demonstrating that our filtering process induces no spurious peak and that the radial velocity signal we detect is unrelated to activity.

A careful re-analysis of our original data¹² demonstrates that, despite being affected by intrinsic variability from the host star, they nonetheless confirm the presence of the planet signal detected in our new data. In particular, applying our filtering analysis to the main subset of our original data (featuring dense enough coverage for our technique to perform reliably) yields filtered radial velocities agreeing well with those derived from the new data. Fitting both sets together further improves the confidence level at which the planet is detected (see Extended Data Fig. 2c).

The detection we report among the small sample of weak-line T Tauri stars (~ 10) already studied with MaTYSSSE suggests that close-in giant planets are potentially more frequent around T Tauri stars than around mature low-mass stars, $\sim 1\%$ of which are known to host hot Jupiters^{25,26}. Poisson statistics, however, indicate that there is still a $\sim 10\%$ chance that hot Jupiters are similarly frequent for both populations. A more quantitative conclusion will have to await a thorough analysis of all MaTYSSSE data, complemented by the large T Tauri star survey to be carried out with SPIRou, the new near-infrared cryogenic spectropolarimeter and velocimeter to be installed at CFHT in 2017.

The broad consensus is that hot Jupiters form beyond a few astronomical units from their host stars and migrate inwards to their eventual close-in orbits. The delivery of hot Jupiters to these orbits may result from planet–disk interactions² (disk migration) or from dynamical interactions with a planetary³ or a stellar companion, followed by orbital circularisation (high-eccentricity migration). These highly debated migration channels were proposed as tentative explanations for hot Jupiters with orbits either well aligned or misaligned with

the spin axis of their host stars²⁷. Our detection of a hot Jupiter on a ~ 5 -day circular (or moderately eccentric) orbit around a 2-Myr-old star is most naturally explained by the mechanism of hot Jupiter delivery by disk migration, which produces hot Jupiters with nearly circular orbits, rather than by planet–planet scattering, which generates highly eccentric ($e > 0.9$) hot Jupiters whose circularization timescales are at least 100 to 1,000 times longer²⁸ than the age of V830 Tau given typical tidal dissipation factors in giant planets²⁹. Our result thus yields strong support to the theory of giant planet migration in gaseous protoplanetary disks² and confirms that the architecture of planetary systems, on which hot Jupiters have a strong impact, is probably dynamic from the very early stages of planetary formation.

Global models of planet formation and evolution show that giant planets can reach a mass and an orbital period similar to those of V830 Tau b in 2–3 Myr, whether formation occurs through core accretion or disk gravitational instability. Large uncertainties in these models make it difficult to accurately predict the occurrence rate of hot Jupiters and thus to associate V830 Tau b with either formation scenario. The ~ 300 G dipole of the star's magnetic field¹² is strong enough to have disrupted the central 0.06 AU of the now-dissipated disk for accretion rates $< 2 \times 10^{-10} M_{\odot} \text{ yr}^{-1}$. For rates of $\sim 10^{-9} M_{\odot} \text{ yr}^{-1}$, more typical of those of the classical T Tauri stars that still feed from their disks, a > 700 G dipole field is required, which is again compatible with observations of classical T Tauri stars similar to V830 Tau^{12,30}. This shows that the field of V830 Tau may well have stopped the planet within the magnetospheric gap¹ at the end of its disk migration, and saved it from falling into the star.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 February; accepted 28 April 2016.

Published online 20 June 2016.

- Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature* **380**, 606–607 (1996).
- Baruteau, C. et al. in *Protostars and Planets VI* (eds Beuther, H., Klessen, R., Dullemond, C. & Henning, T.) Vol. 914, 667–689 (Univ. Arizona Press, 2014).
- Chatterjee, S., Ford, E. B., Matsumura, S. & Rasio, F. A. Dynamical outcomes of planet–planet scattering. *Astrophys. J.* **686**, 580–602 (2008).
- Setiawan, J. et al. A young massive planet in a star–disk system. *Nature* **451**, 38–41 (2008).
- Huélamo, N. et al. TW Hydrae: evidence of stellar spots instead of a hot Jupiter. *Astron. Astrophys.* **489**, L9–L13 (2008).
- van Eyken, J. C. et al. The PTF Orion project: a possible planet transiting a T-Tauri star. *Astrophys. J.* **755**, 42–55 (2012).
- Kraus, A. L. & Ireland, M. J. LkCa 15: a young exoplanet caught at formation? *Astrophys. J.* **745**, 5–16 (2012).
- Sallum, S. et al. Accreting protoplanets in the LkCa 15 transition disk. *Nature* **527**, 342–344 (2015).
- Donati, J.-F. et al. The large-scale magnetic field and poleward mass accretion of the classical T Tauri star TW Hya. *Mon. Not. R. Astron. Soc.* **417**, 472–487 (2011).
- Mahmud, N. I. et al. Starspot-induced optical and infrared radial velocity variability in T Tauri star Hubble I 4. *Astrophys. J.* **736**, 123–131 (2011).
- Donati, J.-F. et al. Modeling the magnetic activity and filtering radial velocity curves of young Suns: the weak-line T Tauri star LkCa 4. *Mon. Not. R. Astron. Soc.* **444**, 3220–3229 (2014).
- Donati, J.-F. et al. Magnetic activity and hot Jupiters of young Suns: the weak-line T Tauri stars V819 Tau and V830 Tau. *Mon. Not. R. Astron. Soc.* **453**, 3707–3720 (2015).
- Grankin, K. N., Bouvier, J., Herbst, W. & Melnikov, S. Y. Results of the ROTOR-program. II. The long-term photometric variability of weak-line T Tauri stars. *Astron. Astrophys.* **479**, 827–843 (2008).
- Siess, L., Dufour, E. & Forestini, M. An internet server for pre-main sequence tracks of low- and intermediate-mass stars. *Astron. Astrophys.* **358**, 593–599 (2000).
- Kraus, A. L., Ireland, M. J., Hillenbrand, L. A. & Martinache, F. The role of multiplicity in disk evolution and planet formation. *Astrophys. J.* **745**, 19–29 (2012).
- Donati, J.-F., Semel, M., Carter, B. D., Rees, D. E. & Collier Cameron, A. Spectropolarimetric observations of active stars. *Mon. Not. R. Astron. Soc.* **291**, 658–682 (1997).
- Vogt, S. S., Penrod, G. D. & Hatzes, A. P. Doppler images of rotating stars using maximum entropy image reconstruction. *Astrophys. J.* **321**, 496–515 (1987).
- Donati, J.-F. & Collier Cameron, A. Differential rotation and magnetic polarity patterns on AB Doradus. *Mon. Not. R. Astron. Soc.* **291**, 1–19 (1997).

19. Donati, J.-F., Collier Cameron, A. & Petit, P. Temporal fluctuations in the differential rotation of cool active stars. *Mon. Not. R. Astron. Soc.* **345**, 1187–1199 (2003).
20. Morin, J. *et al.* The stable magnetic field of the fully convective star V374 Peg. *Mon. Not. R. Astron. Soc.* **384**, 77–86 (2008).
21. Mortier, A., Faria, J. P., Correia, C. M., Santerne, A. & Santos, N. C. BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram. *Astron. Astrophys.* **573**, 101–106 (2015).
22. Haywood, R. D. *et al.* The Sun as a planet-host star: proxies from SDO images for HARPS radial-velocity variations. *Mon. Not. R. Astron. Soc.* **457**, 3637–3651 (2016).
23. Lucy, L. B. & Sweeney, M. A. Spectroscopic binaries with circular orbits. *Astron. J.* **76**, 544–556 (1971).
24. Petit, P. *et al.* A maximum entropy approach to detect close-in giant planets around active stars. *Astron. Astrophys.* **584**, 84–91 (2015).
25. Mayor, M. *et al.* The HARPS search for southern extra-solar planets. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets. Preprint at <http://arxiv.org/abs/1109.2497> (2011).
26. Wright, J. T. *et al.* The frequency of hot Jupiters orbiting nearby solar-type stars. *Astrophys. J.* **753**, 160–164 (2012).
27. Winn, J. N. & Fabrycky, D. C. The occurrence and architecture of exoplanetary systems. *Annu. Rev. Astron. Astrophys.* **53**, 409–447 (2015).
28. Ogilvie, G. I. Tidal dissipation in stars and giant planets. *Annu. Rev. Astron. Astrophys.* **52**, 171–210 (2014).
29. Goldreich, P. & Soter, S. *Q* in the Solar System. *Icarus* **5**, 375–389 (1966).
30. Donati, J.-F. *et al.* Magnetometry of the cTTS GQ Lup: non-stationary dynamos and spin evolution of young Suns. *Mon. Not. R. Astron. Soc.* **425**, 2948–2963 (2012).

Acknowledgements This paper is based on observations obtained at the CFHT Corporation (operated by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la

Recherche Scientifique (INSU/CNRS) of France and the University of Hawaii), at the Télescope Bernard Lyot (TBL) (operated by the Observatoire Midi-Pyrénées and by INSU/CNRS), and at the Gemini Observatory (operated by the Association of Universities for Research in Astronomy under a cooperative agreement with the National Science Foundation (NSF) on behalf of the Gemini partnership: the NSF of the United States, the National Research Council of Canada, CONICYT of Chile, the Ministerio de Ciencia, Tecnología e Innovación Productiva of Argentina, and the Ministério da Ciência, Tecnologia e Inovação of Brazil). We thank the Queue Service Observing teams of CFHT, TBL and the Gemini Observatory, without whom this study would not have been possible. We also thank the IDEX initiative of Université Fédérale Toulouse Midi-Pyrénées for awarding a 'Chaire d'Attractivité' to G.H., in the framework of which this work was done. S.A. acknowledges financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo a Pesquisa do Estado de Minas Gerais (Fapemig).

Author Contributions This work merged data collected with two different instruments and three different telescopes. J.F.D. led the data processing, analysis and manuscript preparation. C.M., L.M., L.Y. and E.H. participated in the data collection and data analysis. C.B. contributed to the theoretical implications of the results. All authors, including G.H., S.A., M.T., F.M., J.B., P.P., R.D. and A.C.C., were involved in elaborating the observing proposals, in discussing the results at various stages of the analysis, and in providing contributions to older versions of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F.D. (jean-francois.donati@irap.omp.eu).

METHODS

Spectropolarimetry with ESPaDOnS/NARVAL and the MaTYSSe programme. ESPaDOnS³¹ and NARVAL are twin spectropolarimeters installed at the Cassegrain focus of the 3.6-m CFHT on top of Maunakea (Hawaii), and of the 2-m Bernard Lyot Telescope on top of Pic du Midi (France), respectively. Both include a fibre-fed bench-mounted high-resolution spectrograph, yielding full coverage of the 370–1,000 nm wavelength range in a single exposure at a spectral resolving power of 65,000. ESPaDOnS can also be fed from the 8-m Gemini-N Telescope next to CFHT, through a 300-m fibre link called GRACES³², yielding spectra with either similar resolving power (in star-only mode) or half of it (in star + sky mode). In this run, we secured 16 spectra with ESPaDOnS from 2015 November 17 to December 02 in fair-weather conditions, 16 spectra with NARVAL from 2015 November 10 to December 17 in moderate to good weather, and 16 spectra with ESPaDOnS/GRACES on four different nights towards the end of the run (with 2 spectra in star-only mode and 14 spectra in star + sky mode). The full journal of observations is given in Extended Data Table 1. ESPaDOnS and NARVAL observations were secured in spectropolarimetric mode (circular polarization) in the framework of the MaTYSSe (Magnetic Topologies of Young Stars and the Survival of close-in giant Exoplanets)¹¹ Large Programme, whereas ESPaDOnS/GRACES observations were collected in Director Discretionary Time in spectroscopic mode (no polarimetric unit on Gemini-N). All spectra were derived from raw frames with the reference pipeline implementing optimal extraction and radial velocity correction from telluric lines¹⁶, yielding a typical root-mean-square radial velocity precision of 30 m s⁻¹ (ref. 33).

Deriving mean line profiles with LSD. LSD¹⁶ is a multiline technique similar to cross-correlation, used to derive line profiles with enhanced signal-to-noise ratio S/N from thousands of spectral lines simultaneously. For this study, the line list we used for LSD is derived from spectrum synthesis through model atmospheres computed assuming local thermodynamic equilibrium³⁴, for atmospheric parameters relevant for V830 Tau¹² (effective temperature of 4,250 K, logarithmic gravity of 4.0 and solar metallicity). Resulting S/N in LSD profiles are in the range 950–1,540 (see the journal of observations in Extended Data Table 1), corresponding to average multiplex gains in S/N of ~ 10 . LSD was also applied to circularly polarized spectra to retrieve average Zeeman signatures and longitudinal field estimates¹⁶.

Doppler imaging of stellar surfaces and the modelling of differential rotation. Doppler imaging is a tomographic technique inspired from medical imaging, with which distributions of brightness features and magnetic fields at the surfaces of rotating stars can be reconstructed from time series of high-resolution spectropolarimetric observations. Doppler imaging is based on the fact that, thanks to the Doppler effect, line profiles of rotating stars can be interpreted as one-dimensional images of stellar surfaces, resolved in the Doppler direction but otherwise blurred. By coupling many such one-dimensional images recorded at different rotation phases, one can reliably reconstruct the parent surface distribution that gives rise to the observed line profiles and rotational modulation. First introduced in the late 1980s¹⁷, Doppler imaging has been extensively used to investigate, with unprecedented accuracy, surface features and magnetic fields in cool stars other than the Sun^{35,36}. Technically speaking, Doppler imaging follows the principles of maximum-entropy image reconstruction, and iteratively looks for the image with lowest information content that fits the data at a given χ^2 level. For more details on the imaging process, our previous MaTYSSe studies^{11,12} give a detailed account of all modelling steps. In this Letter, we carried out Doppler imaging modelling using either unpolarized (Stokes I) spectra only, or both unpolarized and circularly polarized (Stokes V) spectra simultaneously, with identical results regarding filtering performances and the extraction of radial velocity signals.

By looking at how surface maps get twisted as a function of time, Doppler imaging can also estimate the amount of latitudinal differential rotation shearing stellar photospheres^{18,19}. In this study we assume a typical solar-like differential rotation law in which the surface rotation rate varies with latitude θ as $\sin^2\theta$, and depends on two main parameters, the rotation rate at the equator Ω_{eq} and the difference in rotation rate between the equator and the pole, $d\Omega$. Both parameters are derived by looking for the pair that minimizes the χ^2 of the fit to the data (at constant information content in the reconstructed image, see Extended Data Fig. 1), whereas the corresponding error bars are computed from the curvature of the χ^2 paraboloid at its minimum¹⁹. Although helpful to achieve a more accurate description of the activity jitter and a cleaner filtering of raw radial velocity curves (at periods P_{rot} and $P_{\text{rot}}/2$ in particular), differential rotation as weak as that of V830 Tau has little impact on the filtered radial velocities; similar conclusions regarding the planet signal are obtained when assuming that V830 Tau is rotating as a solid body. A similar Doppler-imaging-based technique can be used to diagnose the presence of hot Jupiters around active stars²⁴, with the planet parameters replacing those describing differential rotation. This alternate method yields identical results to those presented here for our data set (see Fig. 4).

Filtering LSD profiles from lunar contamination. Spectra recorded in non-photometric conditions near full-moon epochs are often contaminated by solar light reflected off the moon and diffused by clouds. To filter out this pollution, whose location and width is well known at any given epoch but whose strength we want to determine, we implement a dual-step Doppler imaging process. The first step consists in applying Doppler imaging to the set of original LSD profiles, with scaled-up error bars for all pixels potentially affected by lunar contamination; the strength of the lunar contamination is then measured with a Gaussian fit to the residuals, and subtracted from the polluted LSD profiles. In the second step, conventional Doppler imaging is applied to the set of filtered LSD profiles with original error bars. This dual-step process is found to be very efficient when applied to densely sampled data sets like the one presented here, in which profiles at similar phases but different rotation cycles provide a strong constraint on the strength of lunar pollution. In our data, 6 LSD profiles suffer from a strong pollution (rotation cycles 6.0 to 7.6, see Fig. 1), whereas 9 others are affected at a much weaker level.

If we exclude the 6 strongly moon-polluted profiles (or all 15 moon-polluted profiles) from our data set, the radial velocity signal from V830 Tau b is still clearly detected, though with a lower confidence rate of 99.9% (or 99%), reflecting the poorer temporal coverage and the degraded window function (see Fig. 3a and Extended Data Fig. 2a). This check shows that our decontamination process is successful at restoring the original profile distortions and at retaining the radial velocity content, provided the data set is dense enough.

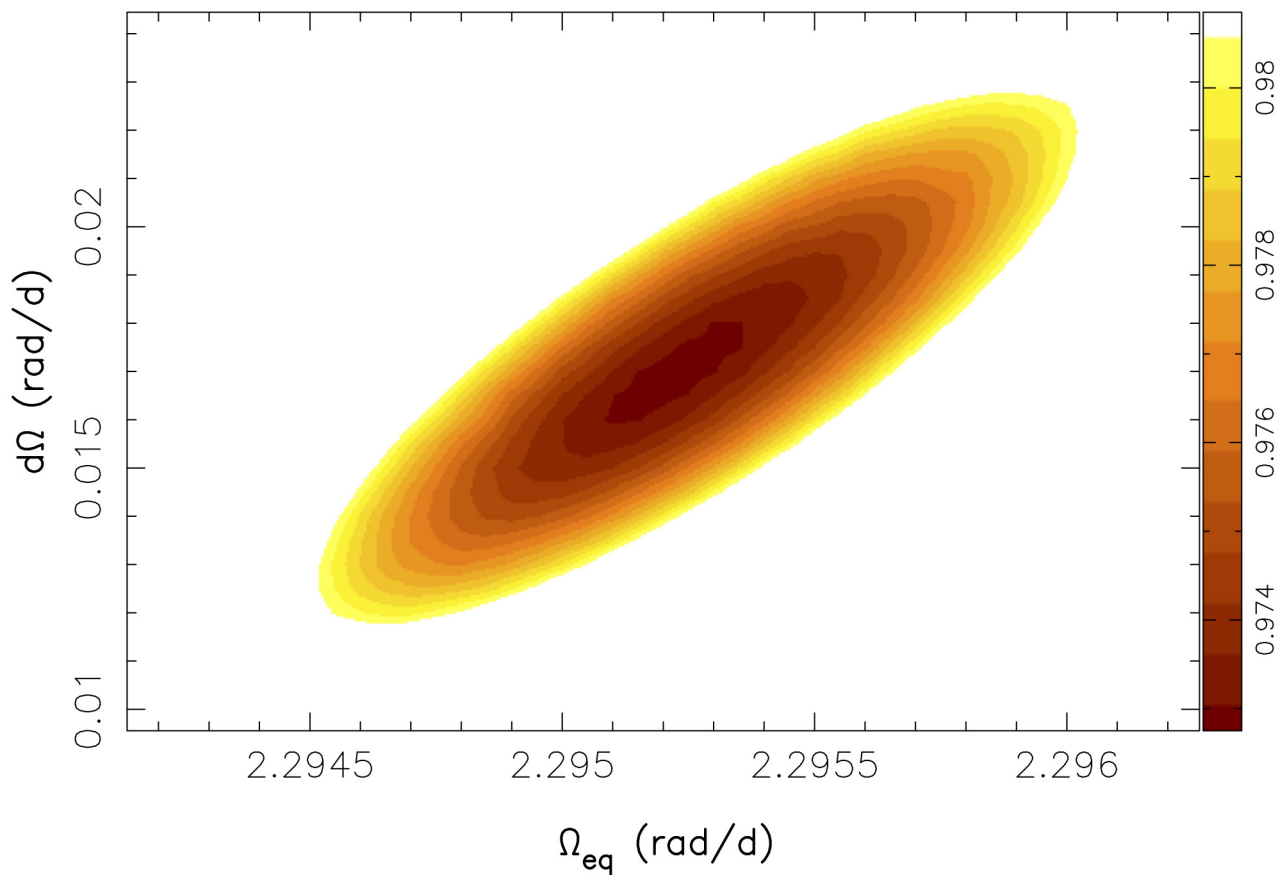
Revisiting the original data set from 2014 December and 2015 January. A careful re-analysis of our original data (consisting of two subsets shifted in time by 17 days¹², see Extended Data Table 3) indicates that variability occurred at the surface of the star between the two subsets. Whereas Doppler imaging succeeds at adjusting the main subset (9 evenly spaced points secured in 2015 January) down to the noise level, as for our new data, fitting both subsets together requires us to lower S/N values in the 2014 December subset by $\sim 15\%$ in order to reach unit reduced χ^2 . This is definite evidence that intrinsic variability (beyond pure differential rotation) occurred at the surface of V830 Tau between both subsets, with profile 6 of the 2014 December subset being the most affected. This variability reflects a modification in the brightness map, subtle enough to affect Doppler imaging no more than moderately, yet large enough to substantially affect filtered radial velocities, which are quite sensitive even to small features in the brightness map. It illustrates how tricky activity filtering can get when dealing with intrinsic variability, and how critical dense and even phase coverage is to diagnose it efficiently.

As a result of this variability, our filtering analysis can only be applied to the individual subsets of our original data, and in fact to no more than the 2015 January subset, the other being far too sparse and uneven for the technique to perform reliably. We find that the filtered radial velocities from the main subset (see Extended Data Table 3) agree with those of our new data; fitting them together improves the confidence level at which the planet is detected, but not the accuracy on the orbital period (see Extended Data Fig. 2c).

Enabling us to shortcut the computation of filtered radial velocities, the Doppler-imaging-based method of adjusting the planet parameters simultaneously with the distribution of surface features²⁴ offers an alternative way to confirm that the radial velocity signal from the detected planet is present in our original data. (This method however still suffers from the inability of Doppler imaging to describe intrinsic variability beyond differential rotation). Freezing the planet orbital period to the value found in our new analysis (4.93 days) and applying this technique to the full set of our original data with only profile 6 removed, we find a semi-amplitude of 67 ± 18 m s⁻¹ for the planet radial velocity signature, which agrees well with the measurement derived from our main study. We also confirm that our original data are better explained by a model including a planet in a 4.93-day orbit than by one with no planet, with a false-alarm probability of $\sim 0.01\%$ (corresponding to a χ^2 increase of 19 for the 644 data points of the fitted LSD profiles).

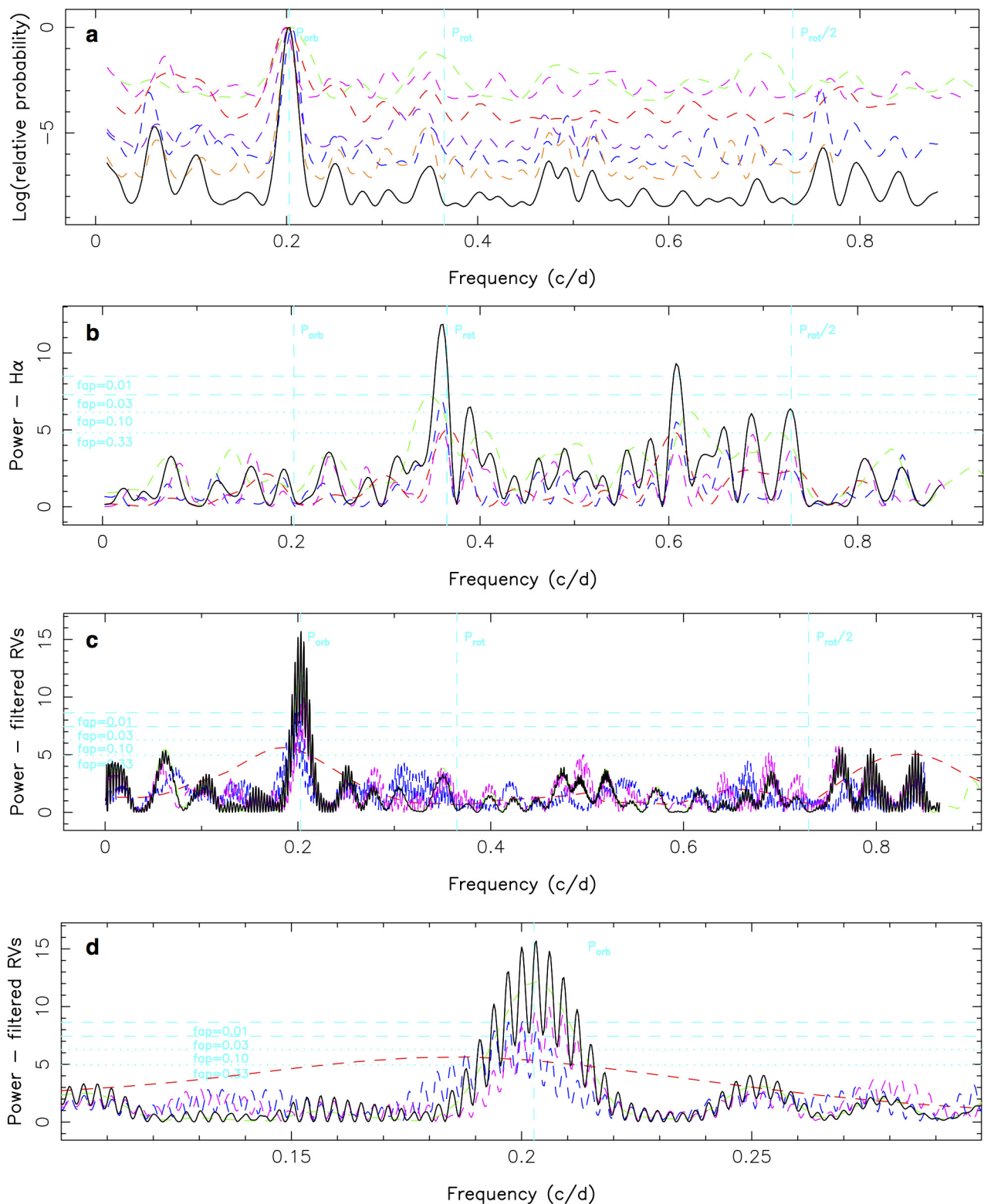
Code availability. The Doppler imaging code used for this study is as yet undocumented and has thus not been released in the public domain.

31. Donati, J.-F. ESPaDOnS: An Echelle SpectroPolarimetric Device for the Observation of Stars at CFHT. *ASP Conf. Proc.* **307**, 41–50 (2003).
32. Chené, A. N. *et al.* GRACES: Gemini remote access to CFHT ESPaDOnS spectrograph through the longest astronomical fiber ever made: experimental phase completed. *Proc. SPIE* **9151**, 47–62 (2014).
33. Moutou, C. *et al.* Spectropolarimetric observations of the transiting planetary system of the K dwarf HD 189733. *Astron. Astrophys.* **473**, 651–660 (2007).
34. Kurucz, R. ATLAS9 atmospheric models, ATLAS9 and SYNTHE routines, spectral line database. CDROM 13 and 18, <http://kurucz.harvard.edu/cdroms.html> (1993).
35. Collier Cameron, A. Spot mapping in cool stars. *Lecture Notes Phys.* **573**, 183–206 (2001).
36. Morin, J. Magnetic fields from low-mass stars to brown dwarfs. *EAS Publ. Ser.* **57**, 165–191 (2012).



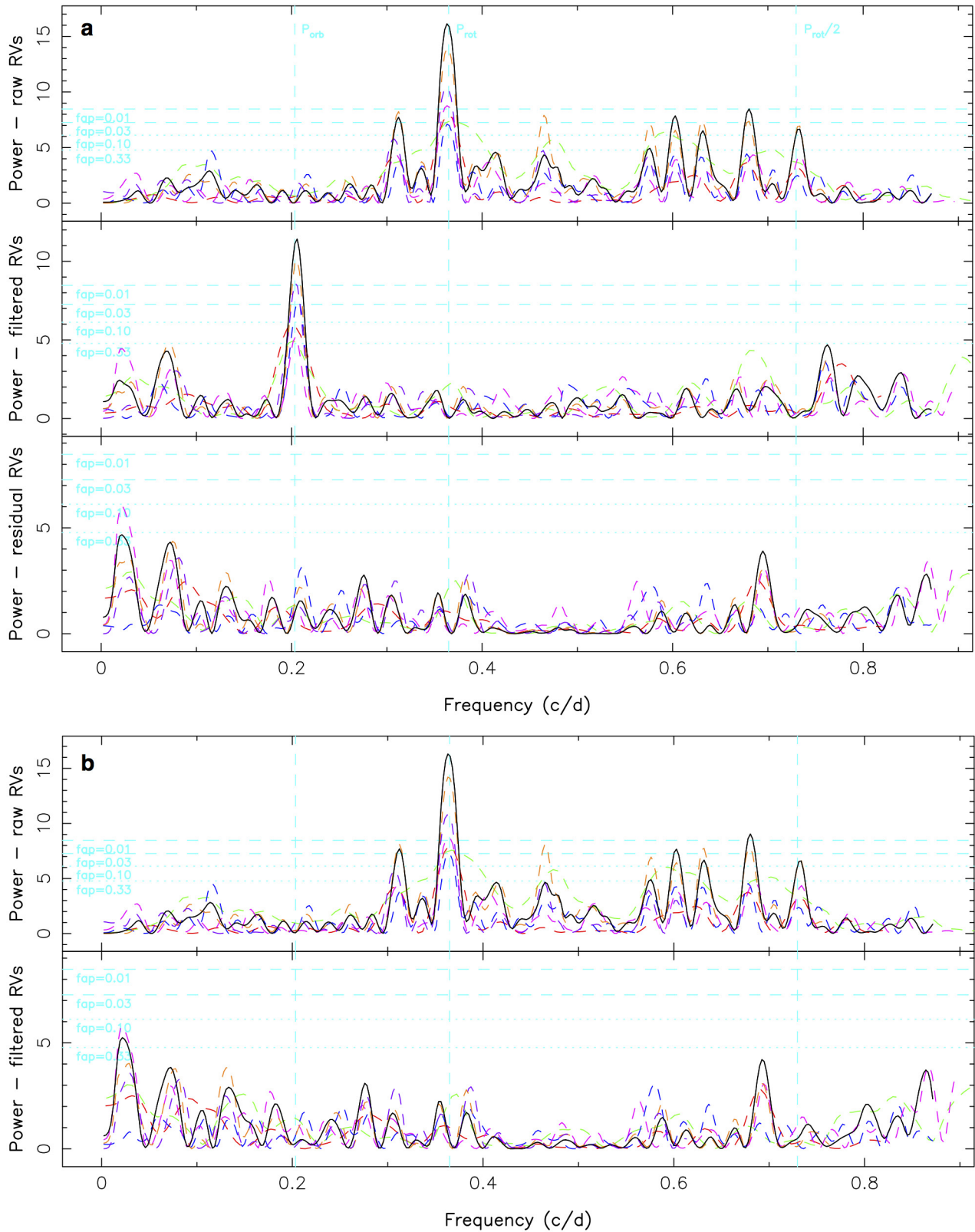
Extended Data Figure 1 | Estimating the surface differential rotation parameter. Variations of the reduced χ^2 as a function of the surface differential rotation parameters Ω_{eq} and $d\Omega$, denoting respectively the rotation rate at the equator and the difference in rotation rate between the equator and the pole (and assuming a solar-like sine-square differential rotation law). The location of the minimum and the local paraboloid

curvature yield the optimal parameters and their respective 1σ error bars¹⁹, equal to 2.29525 ± 0.00020 and 0.0172 ± 0.0014 radians per day. The outer colour contour traces the 99.99% confidence interval (corresponding to a χ^2 increase of 18.4 for a two-parameter optimization problem).



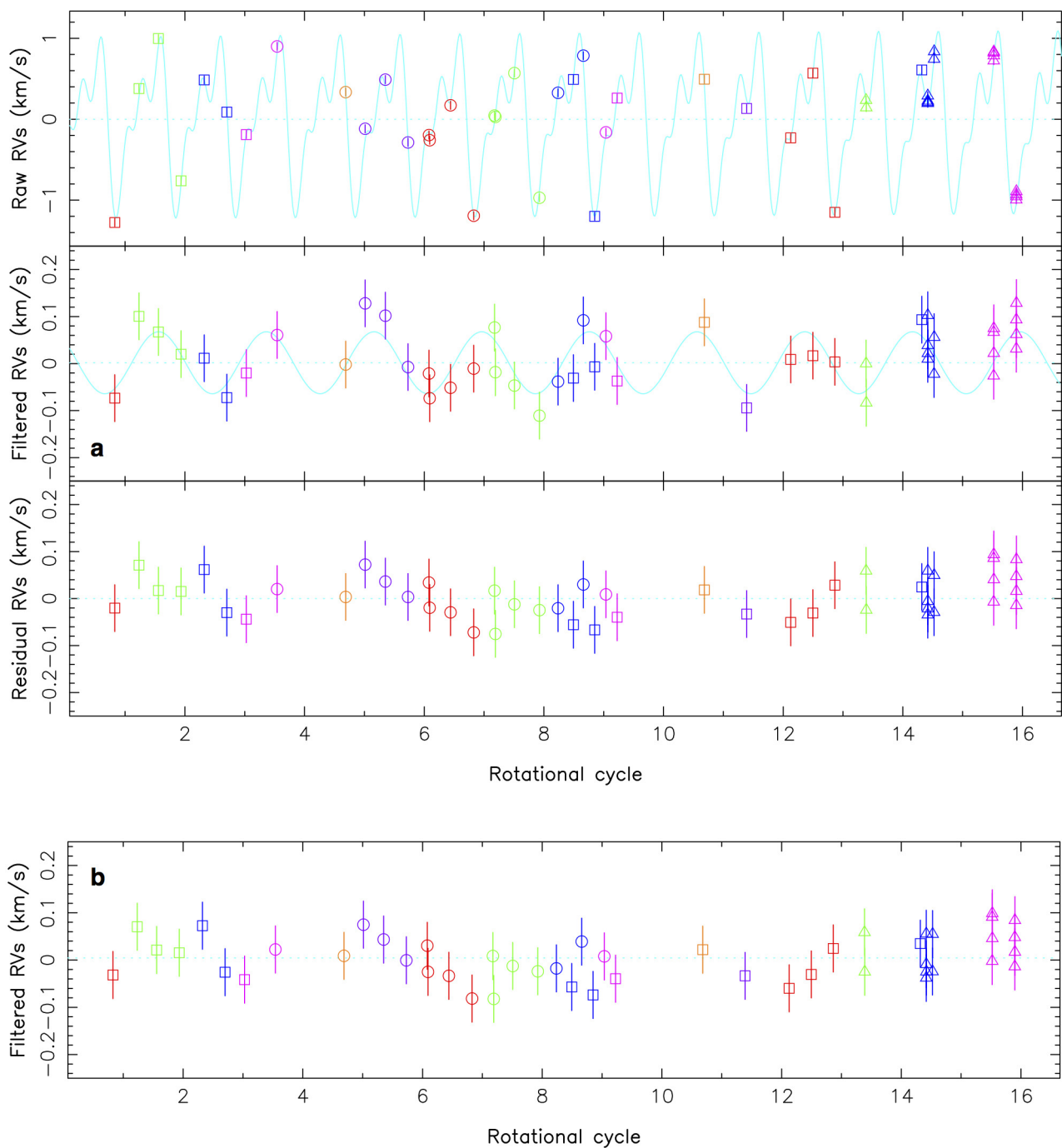
Extended Data Figure 2 | Complementary periodograms. **a**, Same as Fig. 3a (middle plot) using the BGLS approach²¹, showing that the 4.93-day peak we detect is at least 10^5 times more likely than any other features. **b**, Same as Fig. 3b for $H\alpha$ emission, another activity proxy, featuring a clear peak at the stellar rotation period but no power at the planet orbital period. **c**, Same as Fig. 3a (middle plot) for our new data combined with the 2015 January subsample of our original data¹². The planet is now

detected with a higher confidence level ($FAP < 10^{-5}$) but the accuracy of the orbital period is not much improved (with multiple nearby peaks of similar strength). The red and green dashed lines are for the original and new data respectively. **d**, Same as **c**, zooming on the orbital frequency. The orbital period corresponding to the strongest peak is equal to 4.924 ± 0.004 days (1σ error bar).



Extended Data Figure 3 | Periodograms of simulated data. **a**, Same as Fig. 3, for simulated data computed using the brightness map, differential rotation and planet parameters inferred from the real data, and assuming the same coverage and similar S/N (equal for all LSD profiles). As for our observations, the planet signal is detected at a confidence level $>99.9\%$ in the filtered radial velocities despite being invisible in the raw radial velocities, and the planet parameters are well recovered. The periodogram

of the raw radial velocities is very similar to that of Fig. 3, featuring the main peaks (at P_{rot} and $P_{\text{rot}}/2$) and their aliases; residual radial velocities mostly reflect the noise in the data. **b**, Same as the upper two panels of **a** but with no planet included in the simulation. No signal with a confidence level $>90\%$ is recovered in the filtered radial velocities, demonstrating that the filtering process is not generating spurious radial velocity signals, in particular at a period of 4.93 days.



Extended Data Figure 4 | Radial velocity curves of simulated data.

a, Same as Fig. 2a for the simulated data described in Extended Data Fig. 3a (using the same brightness map, differential rotation and planet parameters as those of V830 Tau b). The simulated radial velocities (periodograms shown in Extended Data Fig. 3a) share obvious similarities with the observed ones, and the planet signal is safely recovered

(root-mean-square dispersion of residual radial velocities equal to 45 m s^{-1}). **b**, Same as the middle panel of **a** but with no planet included. As for the periodogram (see Extended Data Fig. 3b, bottom panel), no signal is detected, further illustrating that activity induces no spurious planet signature. As for Fig. 2, simulated radial velocity measurements are depicted in all panels with their 1σ error bars.

Extended Data Table 1 | Journal of observations

UT date (2015)	instrument	BJD (2457300+)	R (K)	t_{exp} (s)	S/N	S/N_{LSD}	rot cycle r (118+)	orb cycle o (5-)	raw RV (km/s)	filt RV (km/s)	RV err (km/s)
Nov 11	NARVAL	37.5066	65	4800	90	1149	0.828	0.332	-1.362	-0.150	0.067
Nov 12	NARVAL	38.6187	65	4800	100	1251	1.233	0.557	0.365	0.074	0.061
Nov 13	NARVAL	39.5077	65	4800	103	1278	1.558	0.738	0.960	0.022	0.059
Nov 14	NARVAL	40.5427	65	4800	95	1233	1.935	0.948	-0.828	-0.035	0.062
Nov 15	NARVAL	41.6051	65	4800	89	1152	2.323	1.163	0.504	0.005	0.067
Nov 16	NARVAL	42.6375	65	4800	105	1257	2.700	1.373	0.137	-0.036	0.060
Nov 17	NARVAL	43.5347	65	4800	84	1113	3.027	1.555	-0.151	0.031	0.070
Nov 18	ESPaDOnS	44.9422	65	2780	162	1420	3.540	1.840	0.830	-0.016	0.048
Nov 21	ESPaDOnS	48.0862	65	2780	168	1380	4.687	2.478	0.382	0.031	0.050
Nov 22	ESPaDOnS	48.9770	65	2780	170	1428	5.012	2.658	-0.199	0.060	0.048
Nov 23	ESPaDOnS	49.9011	65	2780	169	1428	5.350	2.846	0.489	0.087	0.048
Nov 24	ESPaDOnS	50.9411	65	2780	158	1396	5.729	3.057	-0.317	-0.041	0.050
Nov 25	ESPaDOnS	51.9015	65	2780	154	1482	6.079	3.252	-0.312	-0.120	0.048
Nov 25	ESPaDOnS	51.9372	65	2780	164	1480	6.092	3.259	-0.243	-0.040	0.047
Nov 26	ESPaDOnS	52.8917	65	2780	164	1539	6.441	3.452	0.089	-0.126	0.047
Nov 27	ESPaDOnS	53.9535	65	2780	160	1439	6.828	3.668	-1.207	-0.016	0.049
Nov 28	ESPaDOnS	54.9089	65	2780	144	1458	7.177	3.862	-0.014	0.018	0.047
Nov 28	ESPaDOnS	54.9449	65	2780	131	1453	7.190	3.869	0.020	-0.028	0.048
Nov 29	ESPaDOnS	55.8153	65	2780	166	1434	7.507	4.045	0.648	0.029	0.048
Nov 30	ESPaDOnS	56.9632	65	2780	175	1436	7.926	4.278	-0.992	-0.123	0.048
Dec 01	ESPaDOnS	57.8164	65	2780	167	1447	8.237	4.451	0.393	0.014	0.047
Dec 02	NARVAL	58.5216	65	4800	89	1131	8.495	4.594	0.602	0.078	0.069
Dec 02	ESPaDOnS	58.9744	65	2780	168	1428	8.660	4.686	0.823	0.113	0.049
Dec 02	NARVAL	59.4928	65	4800	82	1089	8.849	4.791	-1.054	0.148	0.073
Dec 03	ESPaDOnS	60.0148	65	2780	176	1458	9.039	4.897	-0.163	0.073	0.047
Dec 04	NARVAL	60.5247	65	4800	90	1191	9.225	5.001	0.328	0.019	0.064
Dec 07	NARVAL	64.5183	65	4800	108	1221	10.682	5.811	0.568	0.151	0.064
Dec 09	NARVAL	66.4571	65	4800	89	1161	11.390	6.204	0.155	-0.068	0.066
Dec 11	NARVAL	68.4771	65	4800	88	1218	12.127	6.614	-0.148	0.103	0.064
Dec 12	NARVAL	69.4897	65	4800	69	978	12.496	6.819	0.578	0.024	0.083
Dec 13	NARVAL	70.5011	65	4800	104	1335	12.865	7.024	-1.107	0.055	0.057
Dec 15	ESPaDOnS / GRACES	71.9290	35	540	218	1385	13.386	7.314	0.178	-0.048	0.068
Dec 15	ESPaDOnS / GRACES	71.9359	35	540	214	1388	13.389	7.315	0.148	-0.068	0.068
Dec 17	NARVAL	74.4815	65	4800	90	1257	14.317	7.832	0.617	0.080	0.060
Dec 18	ESPaDOnS / GRACES	74.7548	35	85	72	974	14.417	7.887	0.222	0.058	0.073
Dec 18	ESPaDOnS / GRACES	74.7565	35	85	73	992	14.418	7.888	0.203	0.039	0.072
Dec 18	ESPaDOnS / GRACES	74.7581	35	85	78	990	14.418	7.888	0.222	0.057	0.072
Dec 18	ESPaDOnS / GRACES	74.7597	35	85	77	996	14.419	7.888	0.242	0.078	0.071
Dec 18	ESPaDOnS / GRACES	75.0425	65	360	123	1396	14.522	7.946	0.788	0.016	0.045
Dec 18	ESPaDOnS / GRACES	75.0474	65	360	107	1369	14.524	7.947	0.840	0.053	0.046
Dec 21	ESPaDOnS / GRACES	77.7761	35	85	75	990	15.519	8.500	0.792	0.037	0.072
Dec 21	ESPaDOnS / GRACES	77.7777	35	85	74	986	15.520	8.500	0.766	0.006	0.072
Dec 21	ESPaDOnS / GRACES	77.7793	35	85	74	982	15.520	8.501	0.729	-0.035	0.073
Dec 21	ESPaDOnS / GRACES	77.7810	35	85	74	976	15.521	8.501	0.819	0.050	0.073
Dec 22	ESPaDOnS / GRACES	78.8110	35	85	58	946	15.897	8.710	-0.864	0.161	0.076
Dec 22	ESPaDOnS / GRACES	78.8126	35	85	64	956	15.897	8.710	-0.928	0.093	0.076
Dec 22	ESPaDOnS / GRACES	78.8143	35	85	61	964	15.898	8.711	-0.880	0.139	0.074
Dec 22	ESPaDOnS / GRACES	78.8159	35	85	62	958	15.899	8.711	-0.929	0.087	0.075

Rotation and orbital cycles r and o are respectively given by the ephemeris $\text{BJD} = 2,457,011.8 + 2.741r$ and $2,457,360.52 + 4.93o$. For each observation, R , S/N and S/N_{LSD} list the resolving power, the S/N in the raw spectrum and the S/N in the LSD profile, whereas the last three columns list the raw and filtered radial velocities (RV) and the corresponding 1σ error bars (reflecting both the photon noise and the instrumental radial velocity precision). The first and fifth column respectively list the observing dates in universal time (UT) and the corresponding exposure time t_{exp} .

Extended Data Table 2 | Main parameters of the planet and of the host star

orbital period (d)	K (m/s)	BJD of transit		orbital distance (au)	$m_{\text{planet}} \sin i$ (M_{pl})	m_{planet} (M_{pl})		
4.93±0.05	75±11	2457360.52±0.10		0.057±0.001	0.63±0.11	0.77±0.15		

M_{star} (M_{\odot})	R_{star} (R_{\odot})	age (Myr)	T_{eff} (K)	$\log(L/L_{\odot})$	P_{rot} (d)	$v \sin i$ (km/s)	i (°)	distance (pc)
1.00±0.05	2.0±0.2	~2	4250±50	0.08±0.10	2.741	30.5±0.5	55±10	131±3

Top, main parameters (with 1σ error bars) of V830 Tau b assuming a circular orbit. Bottom, main parameters (with 1σ error bars) of the host star V830 Tau¹², with T_{eff} denoting the effective temperature of the photosphere.

Extended Data Table 3 | Journal of observations for the original data

UT date (2014-2015)	instrument	BJD (2457000+)	R (K)	t_{exp} (s)	S/N	S/N_{LSD}	rot cycle r	raw RV (km/s)	filt RV (km/s)	RV err (km/s)
Dec 20	ESPaDOnS	11.8899	65	2800	170	1501	0.033	0.721		0.049
Dec 21	ESPaDOnS	12.8622	65	2800	170	1504	0.388	0.197		0.050
Dec 22	ESPaDOnS	13.9010	65	2800	180	1520	0.767	-0.513		0.049
Dec 28	ESPaDOnS	20.0190	65	2800	140	1498	2.999	0.446		0.049
Dec 29	ESPaDOnS	20.8759	65	2800	160	1478	3.311	-0.137		0.050
Dec 30	ESPaDOnS	21.8154	65	2800	160	1478	3.654	0.240		0.050
Jan 07	ESPaDOnS	29.8629	65	2800	170	1498	6.590	-0.044	0.031	0.049
Jan 08	ESPaDOnS	30.8217	65	2800	180	1490	6.940	-0.206	-0.031	0.049
Jan 09	ESPaDOnS	31.8181	65	2800	170	1523	7.303	-0.151	-0.062	0.051
Jan 10	ESPaDOnS	32.8186	65	2800	150	1476	7.668	0.044	-0.011	0.050
Jan 11	ESPaDOnS	33.8669	65	2800	180	1495	8.051	1.085	0.076	0.050
Jan 12	ESPaDOnS	34.7215	65	2800	170	1473	8.362	0.246	0.073	0.050
Jan 13	ESPaDOnS	35.7150	65	2800	160	1501	8.725	-0.264	-0.001	0.050
Jan 14	ESPaDOnS	36.7141	65	2800	170	1501	9.089	0.929	-0.067	0.050
Jan 15	ESPaDOnS	37.8043	65	2800	170	1470	9.487	-0.202	-0.052	0.050

Same as Extended Data Table 1 for the 2014 December and 2015 January data^{1,2}. Our filtering technique is found to be reliable only for data sets with dense and regular phase coverage, hence the absence of filtered radial velocities for the 2014 December subset of 2×3 points, which does not satisfy this requirement. Filtered radial velocities from 2015 January 07–15 agree well with those derived from our new data (see Extended Data Table 1).

Tunable two-dimensional arrays of single Rydberg atoms for realizing quantum Ising models

Henning Labuhn^{1*}, Daniel Barredo^{1*}, Sylvain Ravets¹, Sylvain de Léséleuc¹, Tommaso Macrì², Thierry Lahaye¹ & Antoine Browaeys¹

Spin models are the prime example of simplified many-body Hamiltonians used to model complex, strongly correlated real-world materials¹. However, despite the simplified character of such models, their dynamics often cannot be simulated exactly on classical computers when the number of particles exceeds a few tens. For this reason, quantum simulation² of spin Hamiltonians using the tools of atomic and molecular physics has become a very active field over the past years, using ultracold atoms³ or molecules⁴ in optical lattices, or trapped ions⁵. All of these approaches have their own strengths and limitations. Here we report an alternative platform for the study of spin systems, using individual atoms trapped in tunable two-dimensional arrays of optical microtraps with arbitrary geometries, where filling fractions range from 60 to 100 per cent. When excited to high-energy Rydberg D states, the atoms undergo strong interactions whose anisotropic character opens the way to simulating exotic matter⁶. We illustrate the versatility of our system by studying the dynamics of a quantum Ising-like spin-1/2 system in a transverse field with up to 30 spins, for a variety of geometries in one and two dimensions, and for a wide range of interaction strengths. For geometries where the anisotropy is expected to have small effects on the dynamics, we find excellent agreement with *ab initio* simulations of the spin-1/2 system, while for strongly anisotropic situations the multilevel structure of the D states has a measurable influence^{7,8}.

Our findings establish arrays of single Rydberg atoms as a versatile platform for the study of quantum magnetism.

Rydberg atoms have recently attracted a lot of interest for quantum information processing⁹ and quantum simulation¹⁰. In this work, we use individual Rydberg atoms to realize quantum Ising magnets, with unprecedented flexibility in the geometry of the arrays. By shining on the atoms lasers that are resonant with the transition between the ground state $|g\rangle$ and a chosen Rydberg state $|r\rangle$ (Fig. 1a), we implement the Ising-like Hamiltonian

$$H = \sum_i \frac{\hbar\Omega}{2} \sigma_x^i + \sum_{i<j} V_{ij} n^i n^j \quad (1)$$

which acts on the pseudo-spin states $|\downarrow\rangle_i$ and $|\uparrow\rangle_i$ corresponding to states $|g\rangle$ and $|r\rangle$ of atom i , respectively. Here, Ω is the Rabi frequency of the laser coupling, the σ_α^i ($\alpha = x, y, z$) are the Pauli matrices acting on atom i , and $n^i = (1 + \sigma_z^i)/2$ is the number of Rydberg excitations (0 or 1) on site i . The term V_{ij} arises from the van der Waals interaction between atoms i and j when they are both in $|r\rangle$, and scales as $C_6(\theta)|\mathbf{r}_i - \mathbf{r}_j|^{-6}$ with the separation between the atoms $\mathbf{r}_i - \mathbf{r}_j$. Moreover, for $|r\rangle = |nD_{3/2}, m_j = 3/2\rangle$, the van der Waals coefficient C_6 is anisotropic^{7,11}, varying by ~ 3 when the angle θ between the interatomic axis and the quantization axis \hat{z} changes from 0 to $\pi/2$ (Fig. 1a

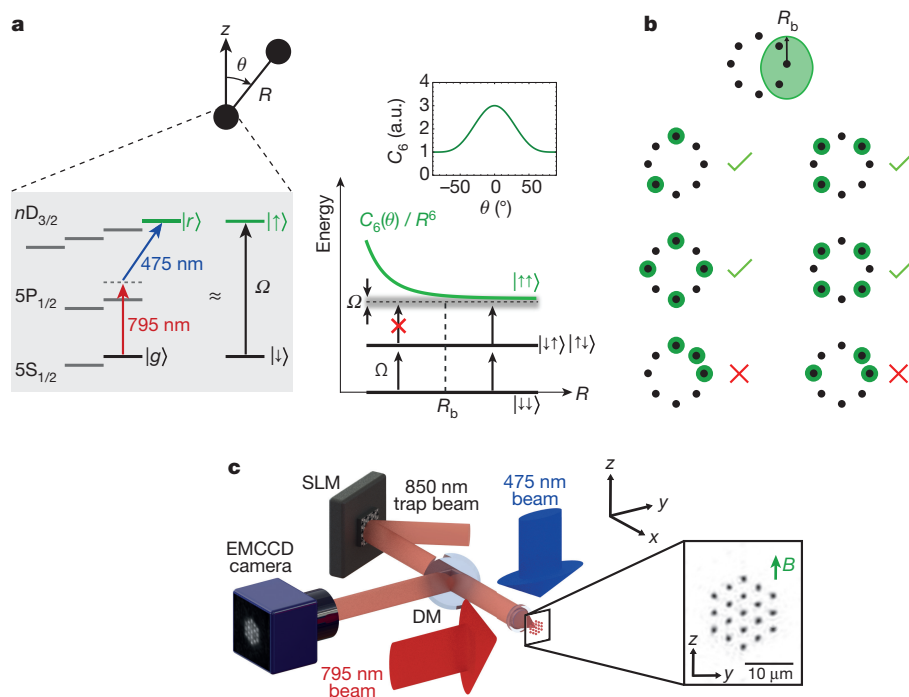


Figure 1 | Many-body dynamics of Rydberg atoms and experimental platform. **a**, Rydberg blockade between two atoms (main panel, lower right), each considered as a two-level system (grey inset): owing to strong interactions between the Rydberg states $|r\rangle$, the excitation of two nearby atoms (within the blockade radius R_b) is inhibited. The use of $nD_{3/2}$ states for $|r\rangle$ gives rise, when the description of the atoms is reduced to a two-level model, to an anisotropic effective van der Waals potential $C_6(\theta)/R^6$ (see inset). **b**, For a value of R_b comparable to the distance between adjacent atoms (top), the dynamics becomes richer. Configurations where two neighbouring atoms are excited are energetically forbidden (red crosses), yielding strong correlations between the Rydberg excitations in the allowed configurations (green ticks). **c**, An array of microtraps is created by imprinting an appropriate phase on a dipole-trap beam. Site-resolved fluorescence of the atoms, at 780 nm, is imaged on an electron-multiplying charge-coupled device (EMCCD) camera using a dichroic mirror (DM). Rydberg excitation beams at 795 and 475 nm are shone onto the atoms. Inset, measured light intensity for an array of $N_t = 19$ traps.

¹Laboratoire Charles Fabry, Institut d'Optique, CNRS, Université Paris Sud 11, 2 avenue Augustin Fresnel, 91127 Palaiseau Cedex, France. ²Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, and International Institute of Physics, Natal-RN, Brazil.

*These authors contributed equally to this work.

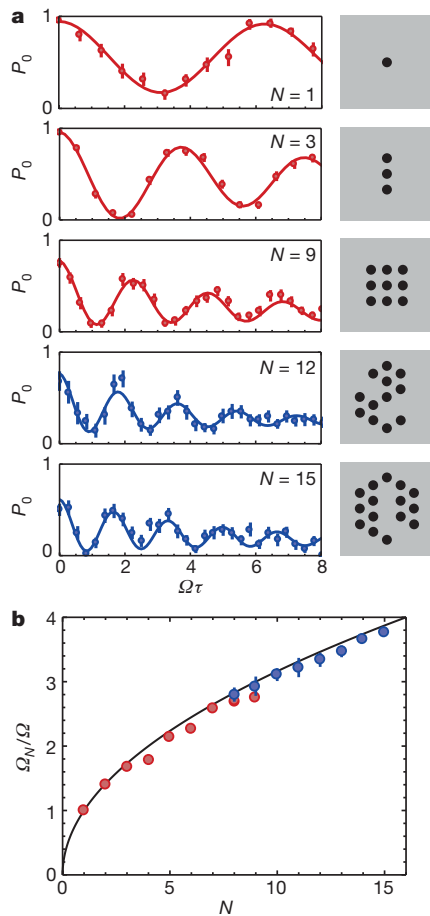


Figure 2 | Collective oscillations in the full Rydberg blockade regime. **a**, Probability P_0 for all N atoms to be in $|g\rangle$ after an excitation pulse of area $\Omega\tau$, for five values of N from 1 to 15. Red points, fully loaded arrays, $n = 82$; blue points, partially loaded triangular arrays of $N_t = 19$ traps, $n = 100$ (error bars show the quantum projection noise for ~ 100 repetitions of the experiment). Solid lines are fits by damped sines of frequency Ω_N . The right panels depict the atomic positions. **b**, Collective oscillation frequency Ω_N/Ω versus N (error bars, sometimes smaller than the symbol size, are s.d.; colour code for points as in **a**). The solid line is the expected \sqrt{N} enhancement.

inset). The strong interactions between the Rydberg states induce correlations in the positions of the excitations (Fig. 1b), as we study experimentally below.

Our set-up (Fig. 1c) has been described in refs 12 and 13. We trap cold ($T \approx 30 \mu\text{K}$) single ^{87}Rb atoms in optical traps with a $1 \mu\text{m}$ waist from a magneto-optical trap (MOT). Using a spatial light modulator (SLM), we create arbitrary, two-dimensional arrays containing $1 \leq N_t \leq 50$ traps, separated by distances $a > 3 \mu\text{m}$. The atomic fluorescence at 780 nm is imaged onto a camera. We observe, in the single-atom regime¹², that the level of fluorescence for each trap alternates randomly between two levels, corresponding to the presence of 0 or 1 atom. The analysis of these N_t fluorescence traces allows us to record, with a time resolution of 50 ms , the current number N of single atoms in the array.

As soon as N exceeds a predefined threshold, we trigger the following experimental sequence. First, the loading of the array is stopped, and a fluorescence image is acquired to record the initial configuration of the atoms, that is, which traps are filled. A 6 G magnetic field is then switched on along the z axis and defines the quantization axis. After initializing all the atoms in $|g\rangle = |5S_{1/2}, F=2, m_F=2\rangle$ by optical pumping, a two-photon Rydberg excitation pulse of duration τ is shone onto the atoms; the Rabi frequency ($\Omega \approx 2\pi \times 1 \text{ MHz}$) is uniform to within 10% over the array. We then acquire a fluorescence image of the final

configuration by switching on the MOT beams. Atoms excited to $|r\rangle$ quickly escape the trapping region, and thus we observe only the atoms that were in $|g\rangle$ after excitation. The atoms that have been lost between the initial and final images are thus assigned to Rydberg states. This detection method has a high efficiency: it only gives a small number of ‘false positives’, as an atom also has a probability $\varepsilon \approx (3 \pm 1)\%$ of being lost, independently of its internal state (Methods).

We first test our system in the conceptually simple situation of fully Rydberg-blockaded ensembles (that is, with at most one Rydberg excitation) containing up to $N = 15$ atoms. Figure 2a shows, for various arrays, the probability P_0 that all N atoms are in $|g\rangle$ at the end of the sequence. We observe high-contrast coherent oscillations, with a frequency enhanced by a factor \sqrt{N} with respect to the single-atom case (Fig. 2b). This characteristic collective oscillation is the hallmark of Rydberg blockade^{14–17}, where multiple excitations are inhibited within a blockaded volume (which, owing to the anisotropy, is close to an ellipsoid, with a major radius R_b defined by $\hbar\Omega = |C_6(0)|/R_b^6$, and a small ‘flattening’ $3^{1/6} \approx 1.2$). This observation is a first step towards the creation of long-lived $|W\rangle$ entangled states (the symmetric combination of the N states with a single excitation) in the ground state⁹.

The fully blockaded regime remains easy to describe theoretically as the blockade naturally truncates the size of the Hilbert space. In contrast, a more challenging regime corresponds to the Rydberg blockade being effective only between nearest neighbours, such that for long enough excitation times, the number of excitations becomes $\sim N/2$. It is therefore desirable to be able to vary the ratio $\alpha = R_b/a$ of the blockade radius to the distance a between sites: for very small or large values of α , the dynamics is simple and the system can easily be compared to numerics, while, for intermediate values of α , the dynamics is challenging to calculate and experimental quantum simulation becomes relevant. Our set-up is particularly adapted to this goal, as we can vary easily both a (reconfiguring the SLM) and R_b (changing the principal quantum number n , we tune C_6 which scales approximately as n^{11}).

This versatility is illustrated in Fig. 3, where we use a fully loaded ring-shaped array of $N = 8$ traps, thus realizing a small spin chain with periodic boundary conditions (PBC). By varying both a and n , we tune the system all the way from independent atoms ($\alpha \ll 1$), where each atom undergoes a Rabi oscillation at frequency Ω , resulting in a Rydberg fraction f_R (defined as the average number of Rydberg excitations divided by N) periodically reaching ~ 1 (Fig. 3a), to a fully blockaded array ($\alpha \gg 1$, Fig. 3c) characterized by collective oscillations at frequency $\sqrt{N}\Omega$ and a maximum $f_R = 1/N$. In between (Fig. 3b, where $\alpha \approx 1.5$), the evolution of $f_R(\tau)$ shows oscillations resulting from the beating of the incommensurate eigenfrequencies of the many-body Hamiltonian, equation (1). Our system allows us to detect the state of each atom, and thus to measure correlation functions. Figure 3d shows the dynamics of the Rydberg–Rydberg pair correlation function:

$$g^{(2)}(k) = \frac{1}{N_t} \sum_i \frac{\langle n_i n_{i+k} \rangle}{\langle n_i \rangle \langle n_{i+k} \rangle} \quad (2)$$

The averaging over all traps does not wash out correlations despite the fact that the system is not fully invariant by translation (Methods). We observe a strong suppression of $g^{(2)}(k)$ for $k = 1$ and $k = 7$, that is, a clear signature of nearest-neighbour blockade. For some times (see for example, $\Omega\tau = 3.1$), we observe an antiferromagnetic-like staggered correlation function, while the average density is uniform (Methods).

The solid lines in all panels of Fig. 3 are obtained by solving the Schrödinger equation governed by equation (1) using the independently measured experimental parameters, and then including the effects of the finite detection errors ε (Methods). One observes an overall agreement with the data, although some small discrepancies can clearly be noticed, especially at longer times. We attribute them to the Zeeman structure of Rydberg D states, which is not taken into account in our modelling by a spin-1/2: for $\theta \neq 0$, the van der Waals interaction

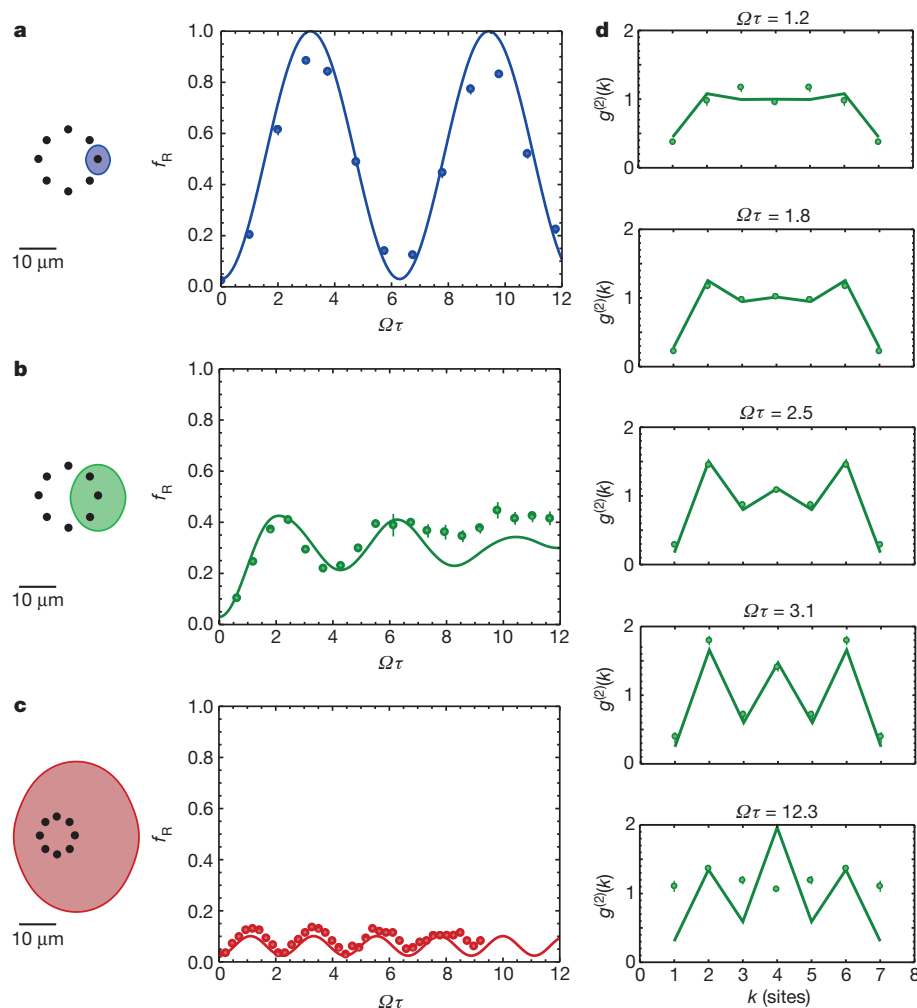


Figure 3 | Tuning interactions in an eight-spin chain with PBC. **a**, Independent atoms ($R_b < a$). Right, the Rydberg fraction f_R oscillates between ~ 0 and ~ 1 with the single-atom Rabi frequency Ω . **b**, Strongly correlated regime ($R_b \approx 1.5a$). Right, the Rydberg fraction shows an oscillatory behaviour involving several frequencies. **c**, Fully blocked regime: f_R oscillates at $\sqrt{N}\Omega$ (right), and reaches a maximum of $1/N$. In **a–c**, the left diagram shows the eight spin chain, and the shaded ellipsoids illustrate the (anisotropic) blockade volume. **d**, The Rydberg–Rydberg pair correlation function, for the parameters of **b**, is shown for increasing values of $\Omega\tau$ (top to bottom). In all plots, the solid lines are obtained by numerically solving the time-dependent Schrödinger equation, and then including detection errors ($\varepsilon = 3\%$). Error bars (often smaller than symbol size) denote s.e.m.

couples $|r\rangle$ to other Zeeman states, leading to a slow increase in the number of excitations (Methods).

We now study two systems containing a larger number of atoms. We first consider a one-dimensional spin chain with PBC comprising $N_t = 30$ traps and partially loaded with $N = 20 \pm 1.5$ atoms (Fig. 4a; we have checked that the 67% filling fraction does not change qualitatively the physics as compared to a perfect filling, see Methods). Its ‘racetrack’ shape was chosen to optimize homogeneity of the Rabi frequency over

the array. We chose parameters such that $\alpha \approx 4.3(1)$. The Rydberg fraction $f_R(\tau)$ shows initial oscillations before reaching a steady state (Fig. 4b) due to the dephasing of the many incommensurate eigenfrequencies of equation (1) for this large value of N . The pair correlation function (shown in Fig. 4c for $\Omega\tau \approx 2.0$) is strongly suppressed for $k < \alpha$, as expected from blockade physics, before oscillating towards the asymptotic value $g^{(2)}(k \gg a) = 1$ (refs 18, 19). A similar liquid-like correlation function has been observed in two dimensions²⁰. The solid

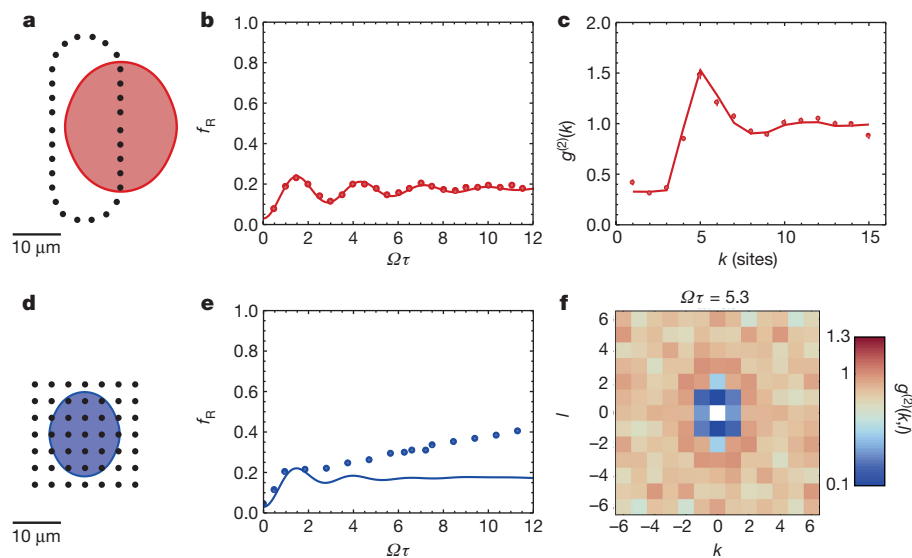


Figure 4 | Ising dynamics in large spin ensembles. **a**, Racetrack-shaped array with $N_t = 30$ traps, loaded with $N = 20 \pm 1.5$ atoms. The blockade radius R_b is about $4.3a$ (shaded ellipsoid). **b**, **c**, Properties of system shown in **a**. **b**, Time evolution of the Rydberg fraction f_R . **c**, Rydberg pair correlation function $g^{(2)}(k)$ for $\Omega\tau \approx 2.0$, showing a strong depletion for $k < R_b$, and contrasted oscillations around the asymptotic value 1 (the data are shown only for $k > 0$, as they are symmetric under the transformation $k \rightarrow -k$). Error bars (most of the time smaller than symbol size) denote the s.e.m. Solid lines are the simulation results without any adjustable parameters. **d**, Square array of 7×7 traps loaded with $N = 28 \pm 1.6$ atoms. The blockade radius is about $2.6a$. **e**, **f**, Properties of system shown in **d**. **e**, Evolution of f_R . Error bars (most of the time smaller than symbol size) denote s.e.m. Solid lines are the simulation results without any adjustable parameters. **f**, Rydberg–Rydberg correlation function $g^{(2)}(k, l)$ for $\Omega\tau = 5.3$.

lines in Fig. 4b,c give the result of a full numerical simulation, without any adjustable parameters. Here the agreement with the spin-1/2 model is excellent, as many atom pairs are aligned along the quantization axis, thus making the effects of the anisotropy small. We included the finite value of ε , which has a strong effect on the pair correlations for $k < \alpha$ as $g^{(2)}(k)$ increases from 0 to $2\varepsilon/f_R$ (Methods).

As a final setting, we use a $N_t = 7 \times 7$ two-dimensional square array (Fig. 4d), loaded with $N = 28 \pm 1.6$ atoms (57% filling), for $\alpha = 2.6$. The dynamics of f_R now appears monotonous (Fig. 4e), without the initial oscillations seen above for smaller systems (Figs 3b and 4b). This suggests that with $N \approx 30$ atoms, the behaviour of the system is already close to the many-body behaviour observed in large ensembles²¹, with a fast initial rise of the Rydberg fraction, before it saturates. The simulation captures the initial rise of f_R well, but does not reproduce the slow increase observed at long times, which we attribute again to multilevel effects (that are indeed expected to be strong in this array where the internuclear axes of many pairs lie at a large θ). Figure 4f shows the two-dimensional Rydberg–Rydberg correlation function

$$g^{(2)}(k, l) = \frac{1}{N_t} \sum_{i,j} \frac{\langle n_{ij} n_{i+k, j+l} \rangle}{\langle n_{ij} \rangle \langle n_{i+k, j+l} \rangle} \quad (3)$$

where n_{ij} refers to the site with coordinates (ia, ja) . Although the system has open boundaries and thus does not show translational invariance, the averaging over the traps in equation (3) does not wash out correlations as R_b is small compared to the system size. We observe a depletion of the correlation function close to the origin due to blockade. The anisotropy of the interaction is visible, as the depletion region is elliptical, with a flattening close to the expected value 1.2. We observe, in the full time evolution of the correlation function (Methods), that blockade volumes become more densely packed with increasing time.

The wide tunability of geometry and interactions demonstrated here opens the way to investigating the physics of spin systems with tens of particles. Our platform, especially when combined with quasi-deterministic loading of optical tweezers as demonstrated recently^{22,23}, will be ideally suited for studying the transition from few- to many-body physics²⁴, thermalization in strongly interacting closed quantum systems²⁵, or the dynamical emergence of entanglement following a quantum quench²⁶. Using resonant dipole–dipole interactions between different Rydberg states²⁷, XY Hamiltonians with long-range couplings²⁸ could also be implemented. Finally, exploiting the Zeeman structure of Rydberg states holds the promise of implementing more complex Hamiltonians, to explore for instance the physics of higher spins²⁹, or to realize topological insulators³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 January; accepted 11 April 2016.

Published online 1 June 2016.

1. Auerbach, A. *Interacting Electrons and Quantum Magnetism* (Springer, New York, 1994).
2. Georgescu, I. M., Ashhab, S. & Nori, F. Quantum simulation. *Rev. Mod. Phys.* **86**, 153–185 (2014).
3. Bloch, I., Dalibard, J. & Nascimbène, S. Quantum simulations with ultracold quantum gases. *Nature Phys.* **8**, 267–276 (2012).
4. Yan, B. *et al.* Observation of dipolar spin-exchange interactions with lattice-confined polar molecules. *Nature* **501**, 521–525 (2013).
5. Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nature Phys.* **8**, 277–284 (2012).

6. Glaetzle, A. W. *et al.* Quantum spin-ice and dimer models with Rydberg atoms. *Phys. Rev. X* **4**, 041037 (2014).
7. Vermersch, B., Glaetzle, A. W. & Zoller, P. Magic distances in the blockade mechanism of Rydberg P and D states. *Phys. Rev. A* **91**, 023411 (2015).
8. Tresp, C. *et al.* Dipolar dephasing of Rydberg D-state polaritons. *Phys. Rev. Lett.* **115**, 083602 (2015).
9. Saffman, S., Walker, T. G. & Mølmer, K. Quantum information with Rydberg atoms. *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
10. Weimer, H., Müller, M., Lesanovsky, I., Zoller, P. & Büchler, H. P. A Rydberg quantum simulator. *Nature Phys.* **6**, 382–388 (2010).
11. Barredo, D. *et al.* Demonstration of a strong Rydberg blockade in three-atom systems with anisotropic interactions. *Phys. Rev. Lett.* **112**, 183002 (2014).
12. Béguin, L., Vernier, A., Chicireanu, R., Lahaye, T. & Browaeys, A. Direct measurement of the van der Waals interaction between two Rydberg atoms. *Phys. Rev. Lett.* **110**, 263201 (2013).
13. Nogrette, F. *et al.* Single-atom trapping in holographic 2D arrays of microtraps with arbitrary geometries. *Phys. Rev. X* **4**, 021034 (2014).
14. Dudin, Y. O., Li, L., Bariani, F. & Kuzmich, A. Observation of coherent many-body Rabi oscillations. *Nature Phys.* **8**, 790–794 (2012).
15. Ebert, M. *et al.* Atomic Fock state preparation using Rydberg blockade. *Phys. Rev. Lett.* **112**, 043602 (2014).
16. Ebert, M., Kwon, M., Walker, T. G. & Saffman, M. Coherence and Rydberg blockade of atomic ensemble qubits. *Phys. Rev. Lett.* **115**, 093601 (2015).
17. Zeiher, J. *et al.* Microscopic characterization of scalable coherent Rydberg superatoms. *Phys. Rev. X* **5**, 031015 (2015).
18. Ates, C. & Lesanovsky, I. Entropic enhancement of spatial correlations in a laser-driven Rydberg gas. *Phys. Rev. A* **86**, 013408 (2012).
19. Petrosyan, D., Höning, M. & Fleischhauer, M. Spatial correlations of Rydberg excitations in optically driven atomic ensembles. *Phys. Rev. A* **87**, 053414 (2013).
20. Schaub, P. *et al.* Observation of spatially ordered structures in a two-dimensional Rydberg gas. *Nature* **491**, 87–91 (2012).
21. Löw, R. *et al.* Universal scaling in a strongly interacting Rydberg gas. *Phys. Rev. A* **80**, 033422 (2009).
22. Lester, B. J., Luick, N., Kaufman, A. M., Reynolds, C. M. & Regal, C. A. Rapid production of uniformly-filled arrays of neutral atoms. *Phys. Rev. Lett.* **115**, 073003 (2015).
23. Fung, Y. H. & Andersen, M. F. Efficient collisional blockade loading of single atom into a tight microtrap. *New J. Phys.* **17**, 073011 (2015).
24. Gaj, A. *et al.* From molecular spectra to a density shift in dense Rydberg gases. *Nature Commun.* **5**, 4546 (2014).
25. Ates, C., Garrahan, J. P. & Lesanovsky, I. Thermalization of a strongly interacting closed spin system: from coherent many-body dynamics to a Fokker-Planck equation. *Phys. Rev. Lett.* **108**, 110603 (2012).
26. Hazzard, K. R. A. *et al.* Quantum correlations and entanglement in far-from-equilibrium spin systems. *Phys. Rev. A* **90**, 063622 (2014).
27. Barredo, D. *et al.* Coherent excitation transfer in a spin chain of three Rydberg atoms. *Phys. Rev. Lett.* **114**, 113002 (2015).
28. Hauke, P. *et al.* Complete devil's staircase and crystal-superfluid transitions in a dipolar XXZ spin chain: a trapped ion quantum simulation. *New J. Phys.* **12**, 113037 (2010).
29. Senko, C. *et al.* Realization of a quantum integer-spin chain with controllable interactions. *Phys. Rev. X* **5**, 021026 (2015).
30. Peter, D. *et al.* Topological bands with a Chern number $C = 2$ by dipolar exchange interactions. *Phys. Rev. A* **91**, 053617 (2015).

Acknowledgements We thank H. Busche for contributions in the early stages of the experiment, I. Lesanovsky, H. P. Büchler and T. Pohl for discussions, and Y. Sortais for a reading of the manuscript. This work benefited from financial support by the EU (FET-Open Xtrack Project HAIRS, H2020 FET-PROACT Project RySQ, and EU Marie-Curie Program ITN COHERENCE FP7-PEOPLE-2010-ITN-265031 (H.L.)), by the ‘PALM’ Labex (project QUANTICA) and by the Région Île-de-France in the framework of DIM Nano-K.

Author Contributions All authors made critical contributions to the work, discussed the results, and contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L. (thierry.lahaye@institutoptique.fr).

Reviewer Information *Nature* thanks C. F. Roos and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Loading of trap arrays. In the single-atom loading regime of optical microtraps, the probability of having a given trap filled with a single atom is $p \approx 1/2$. Therefore, when we monitor the number of loaded traps in view of triggering the experiment (Extended Data Fig. 1a), N fluctuates in time around a mean value $N_t/2$, with fluctuations $\sim \sqrt{N_t}$.

When the number of traps is small, we can impose, as the triggering criterion, waiting until all traps are filled. The average triggering time T_N then increases exponentially with N , as can be seen in Extended Data Fig. 1b. We used this ‘full-loading mode’ for the data of Fig. 1 ($1 \leq N \leq 9$) and Fig. 3 ($N=8$). This exponential scaling sets a practical limit of $N \approx 9$ for fully loaded arrays. For $N=9$, the experimental duty cycle already exceeds one minute.

Because of this, for larger N_t we use partially-loaded arrays. We set the triggering threshold in the tail of the binomial distribution of N , that is, close to $N_t/2 + \sqrt{N_t}$. This allows us to keep a fast repetition rate for the experiment, of the order of 1 s^{-1} , enabling fast data collection. Extended Data Fig. 1c shows the distribution of loaded traps for the ‘racetrack’ array with $N_t=30$ (resp. the $N_t=7 \times 7$ square array), where we set the triggering condition to $N=20$ (resp. $N=30$). Using this triggering procedure, we thus end up with a narrow distribution of atom numbers $N=20 \pm 1.5$ (resp. $N=28 \pm 1.6$), corresponding to a filling fraction of 67% (resp. 57%), significantly above the average $N_t/2$. These strongly subpoissonian distributions of atom numbers are such that the variation in N from experiment to experiment has a negligible effect on the physics studied in Fig. 4; moreover, as for each experiment the initial configuration image is saved, one can if needed post-select experiments where an exact number of atoms was involved (this is how the data in Fig. 2 for $N \geq 10$ were obtained).

Recently, several experiments^{22,23} demonstrated quasi-deterministic loading of single atoms in optical tweezers, reaching $p \approx 90\%$ using modified light-assisted collisions that lead to the loss of only one of the colliding atoms instead of both. A preliminary implementation of these ideas on our set-up gave $p \approx 80\%$ for a single trap. In future work, by using such loading in combination with the real-time triggering based on the measured number of loaded traps, it seems realistic to reach, even in large arrays, filling fractions in excess of 0.9, that is, approaching those obtained in quantum gas microscope experiments using Mott insulators.

Experimental parameters. Extended Data Table 1 summarizes the various values of the parameters of the arrays of traps and of the Rydberg states used for the data presented in the main text, and the resulting values of the dimensionless parameter α . It illustrates the wide tunability offered by the system.

Finite detection errors. Our way to detect that a given atom has been excited to a Rydberg state relies on the fact that we do not detect fluorescence from the corresponding trap in the final configuration image. There is however a small probability ε of losing an atom during the sequence, even if it was in the ground state, thus incorrectly inferring its excitation to a Rydberg state¹¹. These ‘false positive’ detection events affect the measured populations of the N -atom system. One can show that, if P_q is the observed probability of having q Rydberg excitations, and \tilde{P}_p the actual probability of having p Rydberg excitations:

$$P_q = \sum_{p=0}^q \binom{N-p}{q-p} \varepsilon^{q-p} (1-\varepsilon)^{N-q} \tilde{P}_p \quad (4)$$

In principle, one can invert the above linear system relating the observed and actual probabilities³¹, to correct the experimental data for the detection errors. Here we have chosen on the contrary to show the uncorrected populations, and to include detection errors on the theoretical curves instead.

In order to determine the experimental value of ε , we use the initial data points ($\tau=0$) of the data of Fig. 2. Since no Rydberg pulse is sent, we have $\tilde{P}_0=1$, and from equation (4) the observed probability $P_0(\tau=0)$ reads $(1-\varepsilon)^N$. Extended Data Fig. 2a shows the variation of $P_0(0)$ as a function of N , together with a fit which allows us to extract $\varepsilon = (3 \pm 1)\%$, the value we use for the theoretical curves in the main text (see below).

Extended Data Fig. 2b shows the effect of this finite value of ε on the probabilities P_0 , P_1 and P_2 in the full blockade regime, for atom numbers $N=3, 9$ and 15 , clearly illustrating that the ‘false positive’ detection events: first, yield non-zero (and increasing with N) double excitation probabilities (that oscillate in phase with P_1); second, multiply the amplitude of P_0 by a factor $(1-\varepsilon)^N$; and third, reduce the contrast of the P_1 oscillations. Globally, the experimental data (see Extended Data Fig. 3) show these features, superimposed with other imperfections such as damping, not related to the finite value of ε .

Finally, we mention the effect of the detection errors on the correlation functions. In the fully blocked region $k < \alpha$, one ideally expects a vanishing $g^{(2)}$ for $\varepsilon=0$. However, even for a small ε , this value is increased substantially (see, for example, Fig. 4c) to $2\varepsilon/f_R$ where f_R is the Rydberg fraction. Indeed, $g^{(2)}(k=1)$ is given by an average of quantities of the form $\langle n_i n_{i+1} \rangle / (\langle n_i \rangle \langle n_{i+1} \rangle)$. For $\varepsilon=0$, the

numerator vanishes due to blockade; the only possibility of having a non-zero value comes from detection errors. To lowest order in ε , the probability of getting a non-zero value for $n_i n_{i+1}$ is that either atom i is in $|r\rangle$ (probability f_R) and atom $i+1$ is lost (probability ε), or vice versa. This results in a value of $2\varepsilon f_R$ for the numerator, while for the denominator we can use the zeroth-order values $\langle n_i \rangle = \langle n_{i+1} \rangle = f_R$, thus giving $g^{(2)}(1) \approx 2\varepsilon/f_R$, which experimentally can be as large as 0.5.

Additional experimental data. Full Rydberg blockade. Extended Data Fig. 3 shows additional data in the full blockade regime (Fig. 2). In Extended Data Fig. 3a, the arrays of 1 to 9 traps are fully loaded, while in Extended Data Fig. 3b, the 19-trap triangular array is partially loaded with 10 to 15 atoms. In both panels, the left column shows the time evolution of the probability P_0 of recapturing all atoms at the end of the sequence, the middle column shows P_1 , and the right column shows P_2 . The points in Fig. 2a corresponding to $N=8$ and $N=9$ in partially loaded arrays were taken in a similar configuration as for $N=10$ to 15 , but the array contained only $N_t=17$ traps. The curves (not shown here) do not show any noticeable difference with other sets of data. We draw attention to the following.

First, we recognize the effects of the finite detection errors $\varepsilon \neq 0$ on the amplitude and contrast of the collective oscillations discussed above.

Second, the oscillations exhibit some damping, which seems to increase with N . To quantify this, we fit the data by the function

$$P(\tau) = ae^{-\gamma\tau} (\cos^2(\Omega_N\tau/2) + b) + c \quad (5)$$

where a , b , c , γ and Ω_N are adjustable parameters (solid lines). This functional form was chosen to account in a simple way for the asymmetry in the damping. Extended Data Fig. 3c shows the damping rates γ , extracted from the probabilities P_0 as a function of N . We observe an initial increase in the damping rates, which then saturates above $N=5$. An increase with N of the damping rate was observed in other similar blockade experiments^{14–16}. However, even for large number of atoms, the damping rates are small enough that the coherent dynamics dominates over the relevant experimental timescales. We therefore emphasize that in all the other figures of the paper (main text and methods), the theory curves are obtained by disregarding completely any damping, that is, by solving the Schrödinger equation, not a master equation.

Third, we observe that P_2 slowly increases over time for some specific values of N (see in particular $N=4, 6, 9, 13$), corresponding to particular geometries.

We do not have a full understanding of these last two observations, but they may originate from the breaking of the blockade due to the Zeeman structure of the Rydberg states $nD_{3/2}$ (see discussion below).

Eight-atom ring. Extended Data Fig. 4 shows that, within statistical fluctuations, the density of excitations on the eight-atom ring is homogeneous (this remains true at all times), and that the antiferromagnetic-like or crystal-like features obtained for some times, for example, for $\Omega\tau=3.1$, can only be observed in the correlation functions. This illustrates the interest of our set-up, in which spin chains with PBC can be realized easily. On the contrary, in a one-dimensional chain with open boundary conditions, ‘pinning’ of the excitations at specific sites would occur due to edge effects.

Racetrack-shaped array. Extended Data Fig. 5a shows the full evolution of the time correlation function for the data of Fig. 4a–c ($R_b=4.3a$). Extended Data Fig. 5b corresponds to the same settings except for the fact that one now has $R_b=2.4a$.

Square array of 7×7 traps. Extended Data Fig. 6 shows the full time evolution of the two-dimensional Rydberg–Rydberg correlation function $g^{(2)}(k, l)$ for the 7×7 square lattice of Fig. 4d–f. Note that the two-dimensional pair correlation function is calculated using equation (3), which implies that, due to the finite size of the array, the number of terms included in the sum decreases when k, l increase. The normalization takes this variation into account.

Mapping equation (1) on the quantum Ising Hamiltonian. To show the link between Hamiltonian, equation (1), reproduced below,

$$H = \sum_i \frac{\hbar\Omega}{2} \sigma_x^i + \sum_{i<j} V_{ij} n^i n^j \quad (6)$$

and the quantum Ising model, we simply rewrite the operators n^i as $(1 + \sigma_z^i)/2$. Omitting a constant term, we get:

$$H = \sum_i \frac{\hbar\Omega}{2} \sigma_x^i + \sum_i B_i \sigma_z^i + \sum_{i<j} \frac{V_{ij}}{4} \sigma_z^i \sigma_z^j \quad (7)$$

This is the quantum Ising model, with a transverse field proportional to Ω , and a local longitudinal field $B_i = \sum_j V_{ij}/2$ arising from the interactions. In a system with open boundary conditions, this collective longitudinal field is inhomogeneous, and can have observable effects³². In an infinite lattice, or in a system with periodic boundary conditions as realized here in one dimension, the longitudinal field is

homogeneous and could be compensated for by applying a global detuning of the excitation laser.

Anisotropy of the interaction. For a pair of atoms in a $nD_{3/2}$ Rydberg state with the internuclear axis not aligned with the quantization axis, the rigorous description of the van der Waals interaction requires the inclusion of all various Zeeman sublevels; the interaction then takes the form of a 16×16 matrix. To keep the description of a system of N atoms tractable, one can, in the blockade regime, define an effective, anisotropic van der Waals potential⁷ reducing the previous matrix to a single scalar. For $nD_{3/2}$ states, the anisotropy reported in refs 7 and 11 is well reproduced by the simple expression

$$V_{\text{eff}}(r, \theta) = \frac{C_6(0)}{r^6} \left(\frac{1}{3} + \frac{2}{3} \cos^4 \theta \right) \quad (8)$$

with θ the angle between the quantization axis and the internuclear axis, giving a reduction by a factor of three in interaction strength when θ goes from 0 to $\pi/2$ (see inset of Fig. 1a).

Owing to the anisotropy in equation (8), the shape of the blockade volume centred on a Rydberg atom is also anisotropic. However, because of the r^6 -scaling of the interaction, the surface $r(\theta)$ defined by $V_{\text{eff}}(r, \theta) = \hbar\Omega$ is quite well approximated by a prolate spheroid with an aspect ratio of $3^{1/6} \approx 1.2$. In the Figs 1, 3, 4, the shaded regions depicting the blockade volume have the polar equation $r(\theta) = R_b \left(\frac{1}{3} + \frac{2}{3} \cos^4 \theta \right)^{1/6}$.

Numerical simulation of the dynamics. Our theoretical description of the system is based on the mapping of its dynamics into a pseudo-spin 1/2 model with anisotropic long range interactions. We therefore neglect the rich Zeeman structure of the $nD_{3/2}$ states. The numerical calculations rest on the solution of the Schrödinger equation for the Hamiltonian of equation (1) of the main text in a reduced Hilbert space \mathcal{H} . We first write the wavefunction $|\psi\rangle$ of the system with N atoms in terms of states with a fixed number of Rydberg excitations and ground state atoms, which correspond to the eigenstates of the Hamiltonian with vanishing Rabi frequency Ω (refs 17, 33). Then the truncation procedure is based on two complementary steps: first we define the maximum number of Rydberg excitations N_r^{max} that we include in our basis, then second we eliminate those states which display excitations closer than a fixed distance R_0 . Both N_r^{max} and R_0 are adjusted to ensure the convergence of the dynamics. For small samples (Fig. 3) we performed simulations including all 256 basis states, whereas for the racetrack configurations we typically set R_0 smaller than the lattice constant but include up to $N_r^{\text{max}} = 10$ excitations at most, reducing the dimension of \mathcal{H} from $2^{20} \approx 10^6$ to $\sum_{q=0}^{N_r^{\text{max}}} \binom{20}{q} \approx 6 \times 10^5$. For the 7×7 square array with 30 atoms, we set $R_0 = 1.3a$ (much smaller than the blockade radius $R_b = 2.6a$), thus reducing the dimension of \mathcal{H} to $\sim 3 \times 10^6$ (the full Hilbert space is of dimension $2^{30} \approx 10^9$, and using only the truncation criterion on the number of excitations would reduce it to about 5×10^7 , still intractably large). The Schrödinger equation within the truncated Hilbert space is then solved with a standard split-step method for the two non-commuting parts of the Hamiltonian of equation (1). All these calculations were repeated for several realizations of the loading of the arrays (50 realizations for the squared 7×7 configurations and 200 realizations for the case with fewer traps), taking into account the anisotropic interparticle interaction of equation (8). The comparison with experimental data of the average fraction of excitations $f_R = \sum_{q=0}^N q P_q / N$ is done by including the “false positive” detection events as described by equation (4).

The calculation of the $g^{(2)}(k)$ correlation function in Figs 3d and 4c follows the definition of equation (2). However, in contrast with the calculation of the average fraction of the excitations it is not possible to derive an analytical formula for $g^{(2)}(k)$ to properly take into account the detection efficiency of Rydberg excitations (unless $k < \alpha$ as described in section ‘Finite detection errors’). Therefore we implement a standard Monte Carlo algorithm to perform the average of the correlation function over randomly generated configurations which are weighted in $g^{(2)}(k)$ with the initial (quantum) probability extracted from the real time dynamics of the Schrödinger equation. For example, the state $|r_i r_j\rangle$ which contains $N_r = 2$ Rydberg excitations and amplitude $c_{ij}(t)$ can wrongly be detected as the state $|r_i r_j r_q\rangle$ with probability $P = \varepsilon(1 - \varepsilon)^{N-2}$. If the latter state is generated from our sampling algorithm then its weight in the correlation function corresponds to $|c_{ij}(t)|^2$. Finally we average over several hundred randomly generated configurations to obtain well converged results for the correlation function.

Effect of partial loading of large arrays on the observed dynamics. Using the simulations described above, we explore to what extent the partial loading of our larger arrays may change the observed dynamics as compared to the ideal case of full loading.

Extended Data Fig. 7 shows, for the ‘racetrack’ array of Fig. 4a–c, the results of simulations for the experimentally relevant case of partial loading (solid lines, filling fraction $\eta \approx 0.67$) and for the ideal, full loading case (thin dashed lines):

First, Extended Data Fig. 7a also shows the time evolution of the Rydberg fraction f_R . The dynamics is qualitatively similar in the two situations, with initial oscillations that rapidly get damped owing to the dephasing of the many incommensurate eigenenergies of the Hamiltonian. Quantitatively, the initial oscillations are faster in the fully loaded case: this is expected, as each blockade volume contains $1/\eta$ as many atoms, and thus, due to the scaling of the collective Rabi frequency with the number of atoms in a blockade volume, we expect an enhancement of the oscillation frequency by $\sim \eta^{-1/2} \approx 1.2$, close to what we observe. In the same way, the asymptotic Rydberg fraction when $\tau \rightarrow \infty$ is reduced by a factor close to the expected factor η .

Second, Extended Data Fig. 7b shows the pair correlation function $g^{(2)}(k)$ for $\Omega\tau \approx 2.0$. Here again, the changes are moderate, although the oscillations of the correlation function for $k > \alpha$ would be slightly more contrasted for the fully loaded array.

Simulations for the other large array settings give similar results, allowing us to safely conclude that the partial loading of our largest arrays does not affect significantly the observed dynamics. This conclusion would be different for other types of experiments, for instance the transport of a spin excitation in the case of resonant-dipole–dipole interactions.

Approximate translational invariance. For the one-dimensional configurations of the main text (eight-atom ring of Fig. 3b and racetrack-shaped array of 30 traps of Fig. 4a) we plot the spatially averaged pair correlation function

$$g^{(2)}(k) = \frac{1}{N_t} \sum_i \frac{\langle n_i n_{i+k} \rangle}{\langle n_i \rangle \langle n_{i+k} \rangle} \quad (9)$$

where the subscripts label sites. For a system invariant by translation, all terms in the sum are identical, and the averaging over i simply improves the signal to noise ratio. However, our systems are not translationally invariant, in particular because of the anisotropy of the interaction, and a natural question to address is whether the averaging reduces the contrast of the correlation functions. To answer this question, we have calculated the dynamics of the pair correlation function for the eight-atom ring, taking or not taking into account the anisotropy of the interaction (Extended Data Fig. 8). We observe that the contrast reduction due to averaging is very small, thereby validating our choice to perform it for the data shown in the main text.

Effective loss mechanism arising from anisotropic interactions of D states. The agreement between our measurements and the results of the simulations is not perfect for the largest excitation times, in particular for some settings (for example, for some configurations in the full blockade regime, for the eight-atom ring in the partial blockade regime, and for the 7×7 square array), where we observe a gradual increase in the number of measured Rydberg excitations.

These effects could be qualitatively reproduced if the detection errors ε would increase in time. However, the main reason for these losses is that the microtraps are switched off during the excitation (to avoid inhomogeneous light-shifts), and as they are off for a fixed amount of time (3 μ s), independent of τ , we do not, at first sight, expect ε to increase in time. One could imagine however that the presence of the Rydberg excitation lasers may induce extra loss (due to off-resonant scattering for instance), and in this case one would end up having an ε increasing with τ . We have experimentally ruled out this possibility by measuring the recapture probability when shining the Rydberg excitation lasers, detuned from the Rydberg line by ~ 100 MHz, for the full 3 μ s, without measuring any detrimental effect.

A second possible reason would be the motion of the atoms. Owing to their finite temperature, the atoms move during free flight with a velocity $v \approx 50$ nm μ s^{−1}. Now, strictly speaking, the terms corresponding to the laser coupling in equation (1) of the main text are not $\Omega \sigma_x^i$, but $\Omega e^{i\mathbf{k} \cdot \mathbf{r}_i(t)} \sigma_+^i + \text{h.c.}$, where \mathbf{k} is the sum of the wavevectors of the excitation lasers at 795 and 475 nm, and $\mathbf{r}_i(t)$ the position of atom i . Thus, because of the motion, the phase factors of the couplings become time-dependent, which for example, yields a dephasing of the spin wave corresponding to $|W\rangle$ states. However, a numerical simulation of this effect shows that the induced dephasing rates are negligible for our parameters.

We thus believe that the cause of the observed extra losses lies in the large number of interacting Zeeman sublevels when two atoms are excited to $nD_{3/2}$ states: for $\theta \neq 0$, all 16 pair state Zeeman sublevels are coupled together by the van der Waals interaction. For a large number of atoms, this may lead to an effective loss rate from the targeted $|r\rangle$ states into a quasi-continuum comprising all other (weakly interacting) Zeeman states, and hence to a gradual increase of population of the Rydberg manifold. Qualitatively, this interpretation is corroborated by the fact that the observed increase in the number of excitations depends quite strongly on the array geometry: for instance, the data from the racetrack-shaped array (Fig. 4b), for which a majority of interacting atom pairs are almost aligned along the quantization axis z , are well reproduced by the simulations even at long times, unlike in the case of the eight-atom ring or the 7×7 square array

(Figs 3b and 4e), for which many interacting pairs have their internuclear axes strongly inclined with respect to z .

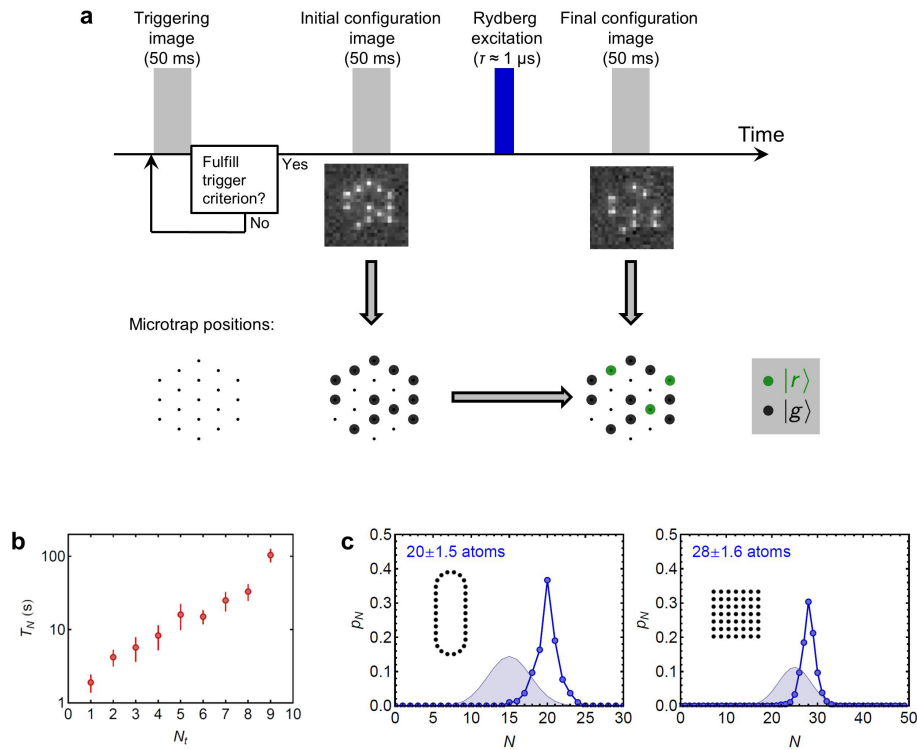
A full simulation of the dynamics of N atoms including the full Zeeman structure of $nD_{3/2}$ states is challenging (neglecting hyperfine structure, the dimension of the Hilbert space is 5^N , as we have 5 states per atom: the ground state and the four Zeeman sublevels of the Rydberg state), and is beyond the scope of this paper. However, in order to test the hypothesis described above, we have simulated the following minimal 'toy model' displaying the effect of a Zeeman structure: we consider a system of $N=6$ atoms in a line, aligned either along the quantization axis z ($\theta=0$) or along y ($\theta=\pi/2$), and excited to the Rydberg state $nP_{1/2}$. The Hilbert space is then of size 3^6 , and we can perform exact diagonalization. The van der Waals Hamiltonian between two atoms (a 4×4 matrix which depends on θ) was taken from equations (4) and (5) of ref. 7.

The red solid lines in Extended Data Fig. 9 show the time evolution of the fraction of Rydberg atoms (whatever the Zeeman state) f_R , calculated using this Hamiltonian with the full Zeeman structure. The black dotted lines show $f_R(t)$ when modelling the system by an assembly of spin-1/2 particles, with an interaction given by the effective potential analogous to equation (8) but calculated for $nP_{1/2}$ states following ref. 7 (for the parameters of the figure, the effective potential is four times as high when $\theta=\pi/2$ as when $\theta=0$). As expected, for $\theta=0$, the two approaches give exactly the same results (Extended Data Fig. 9a). However, for $\theta=\pi/2$, we observe that the two simulations, while giving similar results at short times, disagree significantly at larger times, with the $nP_{1/2}$ structure yielding (as observed in some of our experiments, see for example, Fig. 4e) an increased Rydberg fraction, as the population of the Zeeman state not directly coupled to $|g\rangle$ by the laser (blue dashed line) slowly increases. On the basis of this simulation,

we can thus conclude that in some cases, the anisotropic character of the interaction and the complex Zeeman structure can lead to the observed discrepancy between our data and the modelling by spin-1/2 particles interacting via an effective potential.

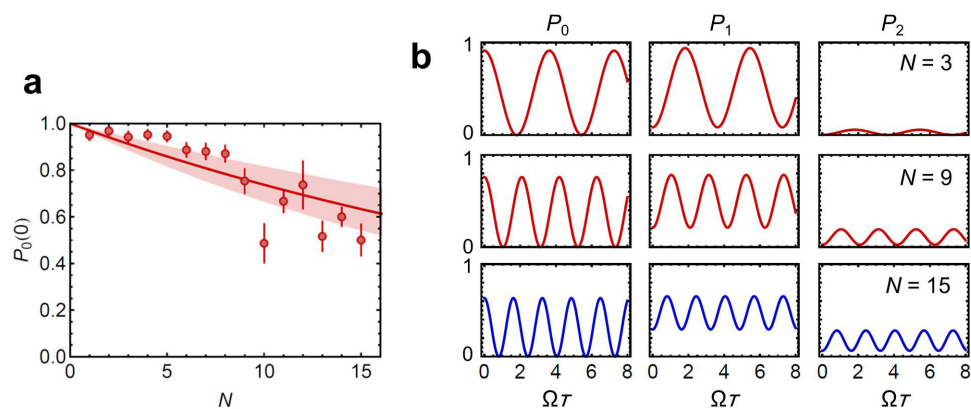
We note however that this behaviour is not universal and that, for a given geometry and a given Rabi frequency, varying the value of n , as well as that of the magnetic field, affects the degree of agreement between the exact dynamics and the simplified spin-1/2 model. In Extended Data Fig. 9, for instance, we chose $n=30$, $B=0.2$ G, and a spacing between the atoms of $a=2$ or $1.6\mu\text{m}$, to display a behaviour qualitatively similar to that of Figs 3b and 4e. This non-universal character suggests that for $nD_{3/2}$ states also, a careful choice of parameters (n , spacing, B field) may allow a very good agreement between the observed dynamics and that of a spin-1/2 model with anisotropic interactions even in geometries where many pairs lie at a large angle from the quantization axis. A comprehensive study of those conditions, along the lines of refs 7, 8, 34, is very important in view of quantum simulation applications, but it is however beyond the scope of this paper and will be the subject of future work.

31. Shen, C. & Duan, L.-M. Correcting detection errors in quantum state engineering through data processing. *New J. Phys.* **14**, 053053 (2012).
32. Zeiher, J. *et al.* Many-body interferometry of a Rydberg-dressed spin lattice. <http://www.arXiv.org/abs/1602.06313> (2016).
33. Schauß, P. *et al.* Crystallization in Ising quantum magnets. *Science* **347**, 1455–1458 (2015).
34. Derevianko, A., Komar, P., Topcu, T., Kroeze, R. M. & Lukin, M. D. Effects of molecular resonances on Rydberg blockade. *Phys. Rev. A* **92**, 063419 (2015).



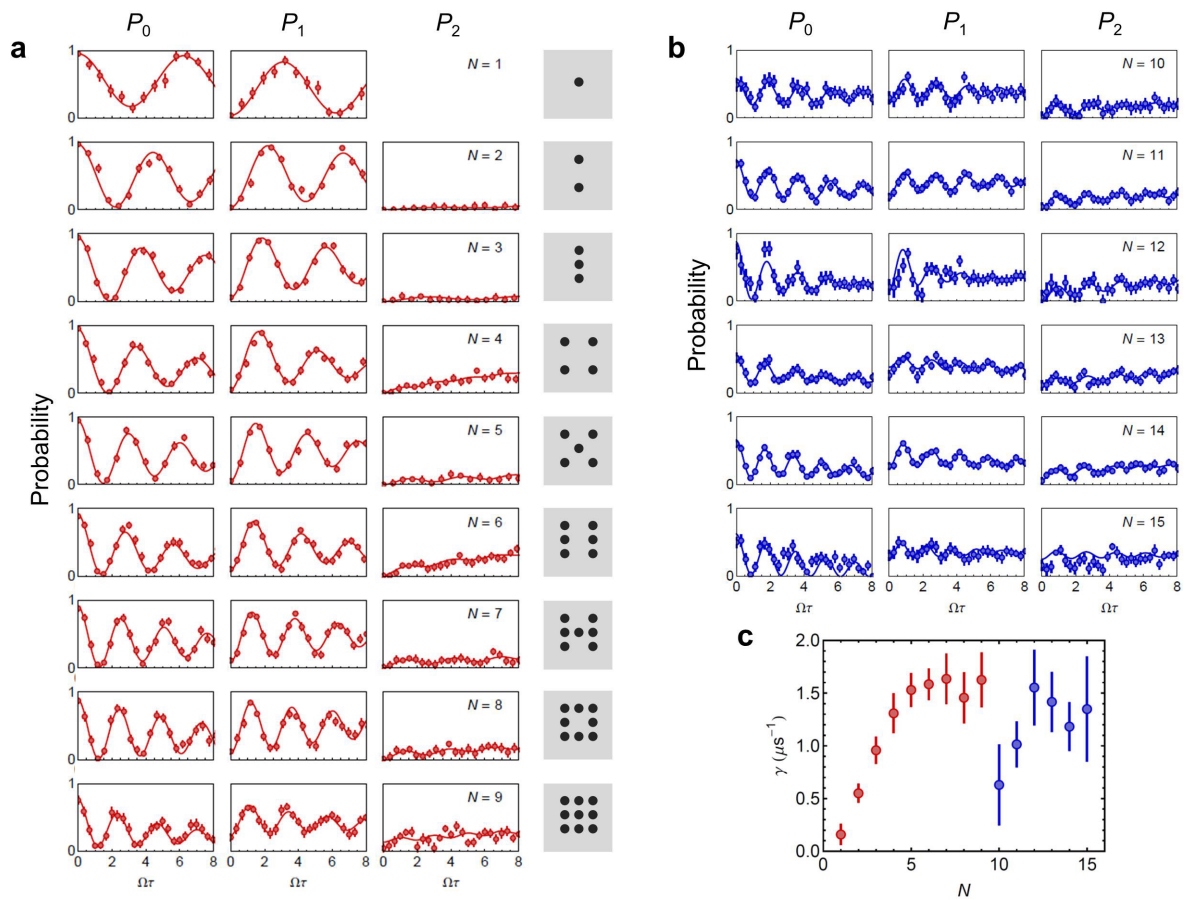
Extended Data Figure 1 | Full and partial loading of arrays. a, Sketch of the experimental sequences. During loading, the camera images are analysed continuously to extract the number of loaded traps. As soon as a triggering criterion is met, the loading is stopped and an image of the initial configuration is acquired. After Rydberg excitation, a final image is acquired, revealing the atoms excited to Rydberg states (green disks at bottom right). **b**, Average triggering time T_N when the triggering criterion is set to $N = N_t$; achieving full loading requires an exponentially long time, limiting the method in practice to $N_t \leq 9$. The triggering times

can vary substantially depending on the density of the magneto-optical trap used to load the array, and the data points shown here correspond to typical conditions used for the data of the main text. Error bars, s.e.m. **c**, Probability p_N of having a number N of loaded traps in the partially loaded regime for the 30-trap 'racetrack' (left) and the 49-trap square array (right; blue dots). The shaded distributions correspond to what would be observed with random triggering. For this partial loading regime, the triggering rate of the experiment is about 1 s^{-1} .



Extended Data Figure 2 | Effect of detection errors. **a**, Experimental determination of ε . From the data of the full blockade experiments (Fig. 2 of main text), we plot the probability P_0 of recapturing all N atoms for $\tau = 0$. The solid line is a fit to the expected dependence $(1 - \varepsilon)^N$, giving

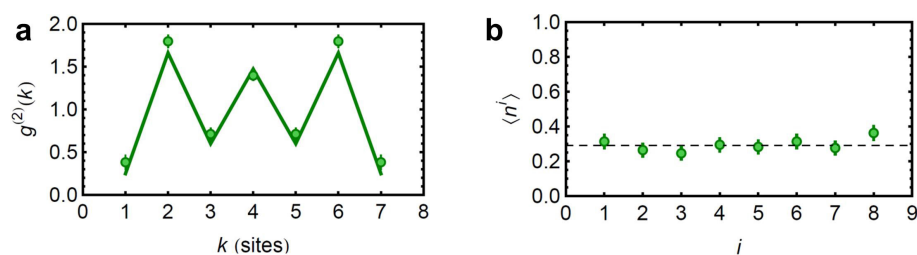
$\varepsilon = 3\%$ (the shaded area corresponds to $2\% < \varepsilon < 4\%$). Error bars, s.e.m. **b**, Calculated probabilities of observing 0, 1 or 2 excitations (columns 1–3) as a function of the excitation pulse area Ωt , assuming a perfect blockade and $\varepsilon = 3\%$, for atom numbers $N = 3, 9, 15$ (rows 1–3).



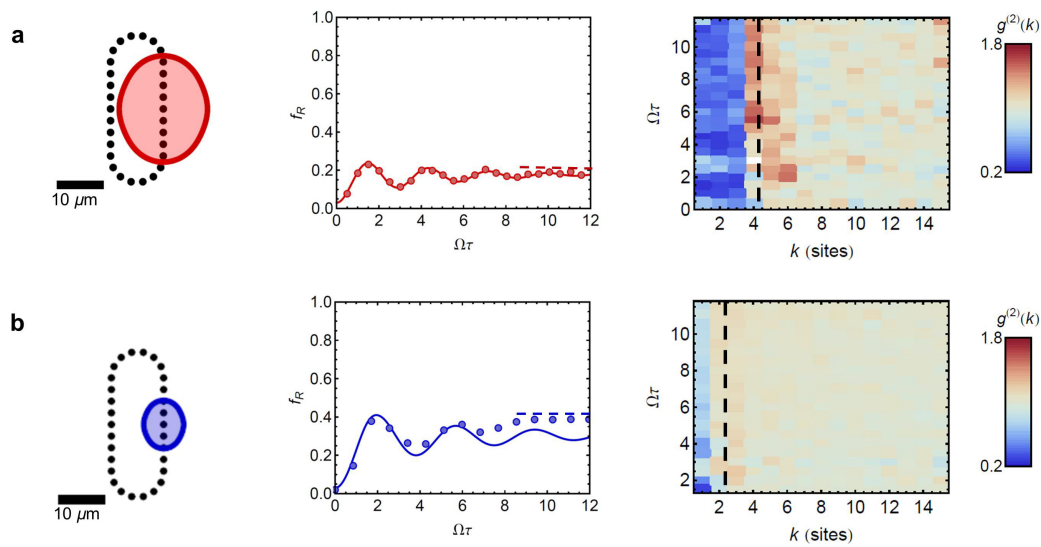
Extended Data Figure 3 | Full data set for the Rydberg blockade data.

a, Fully loaded arrays of 1 to 9 traps ($n = 82$). **b**, Partially loaded array of $N_t = 19$ traps, containing from $N = 10$ to $N = 15$ atoms ($n = 100$). The column on the left shows the probability P_0 of recapturing all atoms, the centre column the probability P_1 of losing just one atom out of N , and

the column on the right the probability P_2 of losing two atoms out of N . The solid lines are fits by equation (5). Error bars, s.e.m. **c**, Damping rate γ extracted from the P_0 data as a function of the number of atoms in the array. Error bars, s.d.

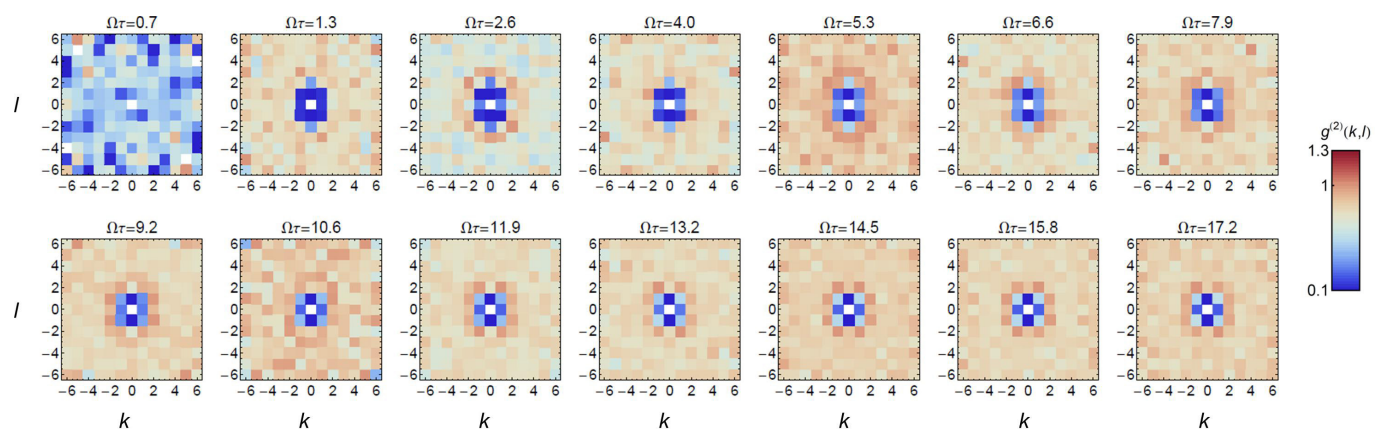


Extended Data Figure 4 | Homogeneous excitation in the eight-atom ring. **a**, For $\Omega\tau = 3.1$, we observe strongly contrasted oscillations in the pair correlation function $g^2(k)$. **b**, The average density of Rydberg excitations, however, is approximately the same on every site. The horizontal dashed line indicates the mean over all sites. Error bars, s.e.m.

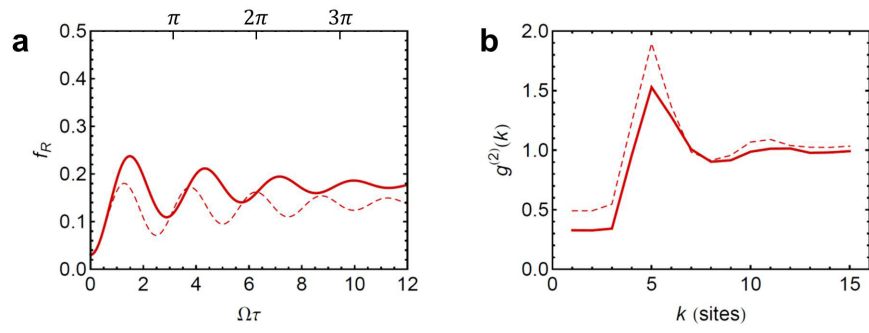


Extended Data Figure 5 | Full time evolution of the correlation functions for the 30-trap, racetrack-shaped chain. **a**, Same as for Fig. 3a–c. The right panel is the time evolution of the pair correlation function, clearly showing that, for times longer than a few Ω^{-1} , the pair correlation function does not evolve significantly. The vertical dashed line indicates the value of the blockade radius. **b**, The principal quantum number is now

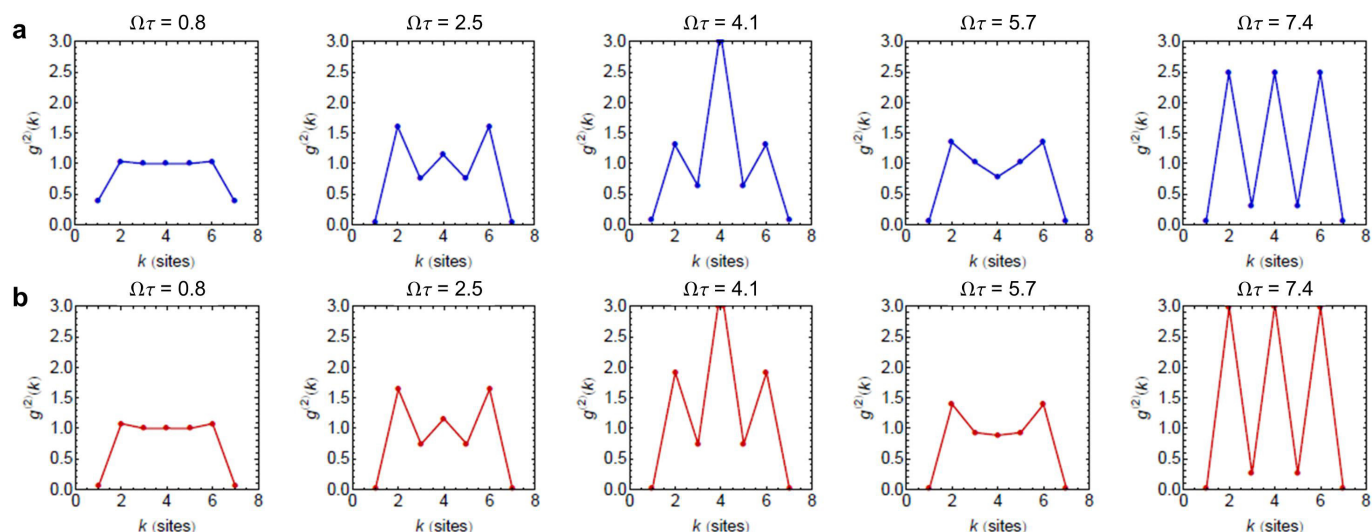
$n = 57$, and the Rabi frequency $\Omega = 2\pi \times 1.7\ \text{MHz}$, such that $R_b = 2.4a$. The central panel shows the time evolution of the Rydberg fraction, and the right panel the time evolution of the pair correlation function. For both **a** and **b**, f_R approaches, at long times, the close-packing limit a/R_b of hard rods of length R_b (dashed horizontal lines).



Extended Data Figure 6 | Full time evolution of the experimental correlation function for the 7×7 square array. One observes the blocked region around $(k, l) = (0, 0)$, with a slight flattening reflecting the anisotropy of the interaction. After a few Ω^{-1} , the correlation function $g^{(2)}(k, l)$ does not evolve any more.

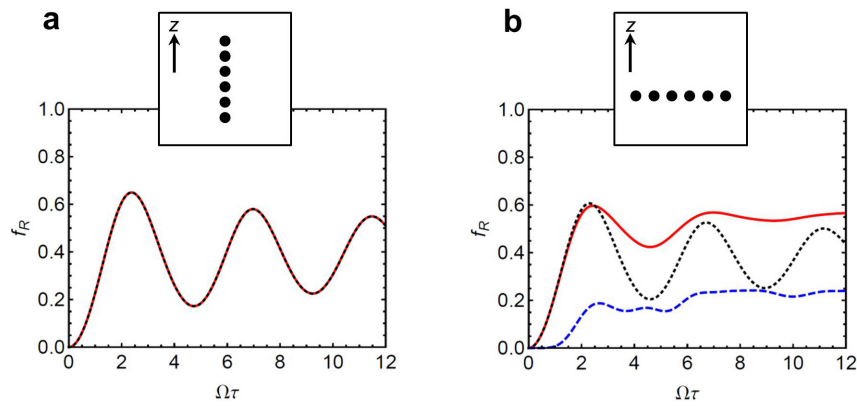


Extended Data Figure 7 | Full versus partial loading for the dynamics and correlations in the case of Fig. 4a–c. **a**, Rydberg fraction as a function of time for the partially loaded (solid line) or fully loaded (thin dashed line) 30-trap array. **b**, Pair correlation function $g^{(2)}(k)$ for $\Omega\tau \approx 2.0$, for the partially loaded (solid line) or fully loaded (thin dashed line) 30-trap array. In both cases, the effect of detection errors ($\varepsilon = 3\%$) is included.



Extended Data Figure 8 | Assessing the validity of the approximation of translational invariance in the eight-atom ring. Calculated pair correlation function $g^{(2)}(k)$ as a function of the excitation time for the eight-atom ring. **a**, Simulation using the experimentally relevant anisotropic interaction, which breaks translational invariance.

b, Simulation with the same parameters as in **a**, except that the angular dependence is neglected (we replace equation (8) by its value for $\theta = 0$), thus re-establishing translational invariance. We observe that the contrast in **a** is reduced, as expected, but only in a marginal way.



Extended Data Figure 9 | Effect of the Zeeman structure of the Rydberg states on the dynamics of the Rydberg fraction f_R . We use the toy model of $nP_{1/2}$ Rydberg states discussed in Methods, with $n = 30$ and $B = 0.2$ G. **a**, The atoms are aligned along the quantization axis z , and spaced by $a = 1.6 \mu\text{m}$ (inset). In this case, the full solution including the Zeeman structure (red solid line) agrees perfectly with the solution of the effective spin-1/2 model with an anisotropic effective potential (as used in all the rest of the paper, black dotted line). **b**, The atoms are aligned

perpendicularly to z , with $a = 2 \mu\text{m}$ (inset; thus keeping the same effective potential interaction between adjacent atoms as in **a**). The full (red solid line) and approximate (black dotted line) solutions agree at short times, but for longer times some population builds up in the other Zeeman sublevel that is not directly coupled to $|g\rangle$ by the laser (blue dashed line), resulting in a slowly increasing excess of Rydberg fraction similar to what is observed experimentally for some configurations.

Extended Data Table 1 | Experimental parameters used for the data presented in the main text

Figure	Trap array parameters				Rydberg state parameters			
	Spacing a (μm)	N_t	N	n	Calculated C_6/h (GHz μm^6)	$\Omega/(2\pi)$ (MHz)	R_b (μm)	α
2a (full)	3.0	1–9	N_t	82	-8.9×10^3	1.5	14	4.5
2a (partial)	3.2	19	10–15	100	-8.0×10^4	1.1	20	6.4
3a	6.3	8	8	54	–6.7	1.6	4.0	0.63
3b	6.3	8	8	61	-7.6×10^2	1.3	9.1	1.4
3c	3.8	8	8	100	-8.0×10^4	0.95	21	5.5
4a,b,c	3.1	30	20 ± 1.5	79	-6.0×10^3	1.0	13.5	4.3
4d,e,f	3.5	49	28 ± 1.6	61	-7.6×10^2	1.4	9.1	2.6

Wide tuning of $\alpha = R_b/a$, over one order of magnitude, is achieved by a combination of changes in a and n (while Ω is kept almost constant).

Synthetic Landau levels for photons

Nathan Schine¹, Albert Ryou¹, Andrey Gromov², Ariel Sommer¹ & Jonathan Simon¹

Synthetic photonic materials are an emerging platform for exploring the interface between microscopic quantum dynamics and macroscopic material properties^{1–5}. Photons experiencing a Lorentz force develop handedness, providing opportunities to study quantum Hall physics and topological quantum science^{6–8}. Here we present an experimental realization of a magnetic field for continuum photons. We trap optical photons in a multimode ring resonator to make a two-dimensional gas of massive bosons, and then employ a non-planar geometry to induce an image rotation on each round-trip⁹. This results in photonic Coriolis/Lorentz and centrifugal forces and so realizes the Fock–Darwin Hamiltonian for photons in a magnetic field and harmonic trap¹⁰. Using spatial- and energy-resolved spectroscopy, we track the resulting photonic eigenstates as radial trapping is reduced, finally observing a photonic Landau level at degeneracy. To circumvent the challenge of trap instability at the centrifugal limit^{10,11}, we constrain the photons to move on a cone. Spectroscopic probes demonstrate flat space (zero curvature) away from the cone tip. At the cone tip, we observe that spatial curvature increases the local density of states, and we measure fractional state number excess consistent with the Wen–Zee theory, providing an experimental test of this theory of electrons in both a magnetic field and curved space^{12–15}. This work opens the door to exploration of the interplay of geometry and topology, and in conjunction with Rydberg electromagnetically induced transparency, enables studies of photonic fractional quantum Hall fluids^{16,17} and direct detection of anyons^{18,19}.

The Lorentz force on a charged particle moving in a magnetic field leads to the unique topological features of quantum Hall systems, including precisely quantized Hall conductance, topologically protected edge transport, and, in the presence of interactions, the predicted anyonic and non-Abelian braiding statistics that form the basis of topological quantum computing²⁰. To controllably explore the emergence of these phenomena, efforts have recently focused on realizing synthetic materials in artificial magnetic fields, and in particular, upon implementations for cold atoms and photons. Successful photonic implementations have employed lattices with engineered tunnelling^{6,21–24}. However, it is desirable to realize artificial magnetic fields in the simpler case of a continuum (lattice-free) material^{7,25,26}, where strong interactions are more easily accessible and the theory maps more directly to fractional quantum Hall systems. In this work, we develop a new approach and demonstrate the first continuum synthetic magnetic field for light.

To achieve photonic Landau levels we harness the powerful analogy between photons in a near-degenerate multimode cavity and massive, trapped 2D particles^{27,28}. Owing to mirror curvature, the transverse dynamics of a running wave resonator are equivalent to those of a 2D quantum harmonic oscillator (Fig. 1a). Non-planar reflections cause the transverse properties of the light field—for example, field profile (image) and polarization vectors—to rotate by an angle ϕ upon a round trip (Fig. 1b). Polarization rotation splits the energy of circularly polarized eigenmodes, while image rotation, in analogy to a rotating frame, introduces Coriolis and centrifugal forces. As the anti-confinement from the rotation compensates the confinement from the mirror

curvature, we are left primarily with a Coriolis force, or equivalently, a Lorentz force. When dynamics are coarse-grained over many round-trips, we arrive at the Fock–Darwin Hamiltonian (see Supplementary Information) $H_{\text{FD}} = \frac{1}{2m} \left(\mathbf{p} - \frac{(qB)^{\text{syn}}}{2} \hat{\mathbf{z}} \times \mathbf{r} \right)^2 + \frac{1}{2} m \omega_{\text{trap}}^2 \mathbf{r}^2$, where m is the dynamical particle mass, \mathbf{p} is the particle's transverse momentum vector, \mathbf{r} is the particle's transverse position vector, $\hat{\mathbf{z}}$ is the longitudinal unit vector, and $\omega_{\text{trap}}/2\pi$ is the (residual) harmonic trapping frequency. The synthetic magnetic field is $\frac{28}{h} \frac{(qB)^{\text{syn}}}{\lambda L_a} = \frac{2\pi}{\lambda L_a} \phi \approx \frac{8\pi}{\lambda L_a} \theta^2$ for small angles θ , where L_a and θ are the on-axis resonator length and opening half-angle (Fig. 1c), and λ is the wavelength of light. When the resonator length is tuned to eliminate residual harmonic trapping, only a Lorentz force remains, and the Hamiltonian describes massive particles in Landau levels, where the n th Landau level has energy $\hbar \omega_c \left(n + \frac{1}{2} \right)$, with ω_c being the cyclotron frequency, and consists of states with angular momentum $l = -n, -n+1, \dots$ in units of the angular momentum quantum \hbar . The synthetic magnetic field is then equivalently given by $(qB)^{\text{syn}}/\hbar = 4/w_0^2$, that is, one flux quantum per area $\pi w_0^2/4$, where w_0 is the resonator $l=0$ mode waist ($1/e^2$ intensity radius). The magnetic length l_B may therefore be identified as $w_0/2$.

Although Landau levels exhibit ‘topological protection’ against localized disorder, long-range potentials may guide the particles to infinity, inducing loss^{11,29}. In our system, the dominant source of long-range disorder is trap asymmetry (astigmatism) that arises from mirror imperfections and off-axis reflection and drives $\Delta l = \pm 2$ transitions (see Supplementary Information). We circumvent this by imposing an additional discrete three-fold rotational symmetry on our Landau levels. To achieve this, we carefully balance transverse and longitudinal energy scales such that only every third angular momentum state is degenerate (see Supplementary Information).

The three-fold symmetry of the Landau levels induces a conical geometry on the 2D space for transverse photon dynamics. To see this, consider a particle which leaves the edge of a particular 120° wedge of the plane; the discrete rotational symmetry requires it to appear on the other side, which is equivalent to wrapping this wedge into a cone (Fig. 1d). Working away from the apex of the cone gives access to flat space Landau levels with every angular momentum state accessible, while working near the apex allows experimental investigation of particle dynamics near a singularity of spatial curvature.

Our experimental resonator consists of four mirrors with nominal radii of curvature $R = (2.5, 5, 5, 2.5)$ cm arranged as shown in Fig. 1c, and has an $l=0$ mode finesse of 2.0×10^4 . The on-axis length $L_a = 1.816$ cm and the opening half-angle $\theta = 16^\circ$ were chosen to create a photonic Landau level while minimizing residual astigmatism. Varying the resonator length by $\sim 20 \mu\text{m}$ adjusts the splitting between states by ~ 1 MHz (see Supplementary Information). Tuning this splitting to zero results in a free spectral range at degeneracy of $\nu_{\text{FSR}} = 3.8209(2)$ GHz. The resonator has an $l=0$ waist size $w_0 = 43 \mu\text{m}$ and a cyclotron frequency $\omega_c = 2\pi \times 2.1671(2)$ GHz, which together yield a photon dynamical mass of $m_{\text{dyn}} = \frac{4\hbar}{\omega_c w_0^2} = 1.84 \times 10^{-5} m_e$, where m_e is the electron mass.

¹Department of Physics and James Franck Institute, University of Chicago, Chicago, Illinois, USA. ²Kadanoff Center for Theoretical Physics, University of Chicago, Chicago, Illinois, USA.

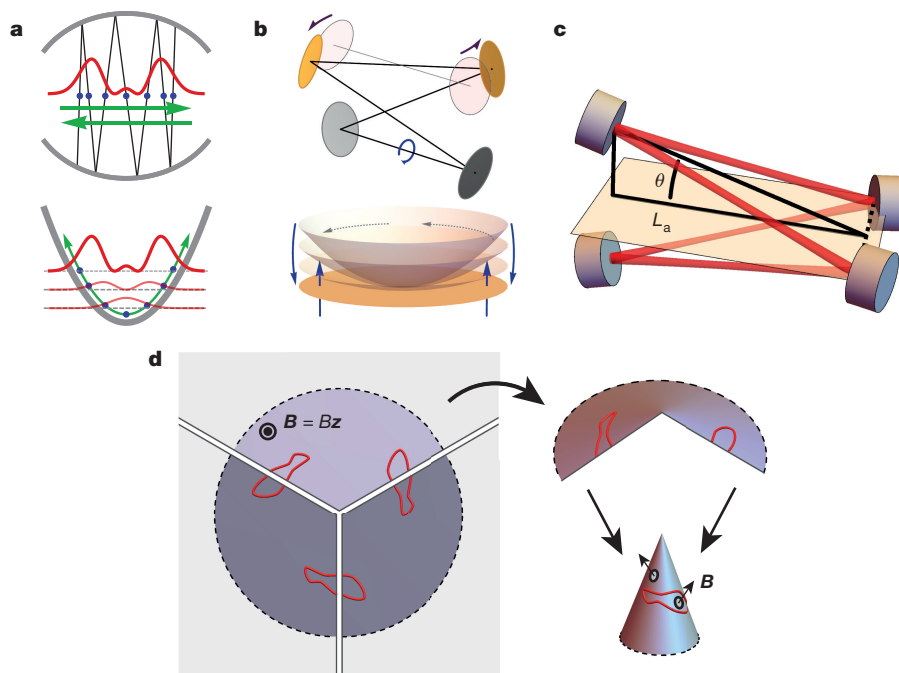


Figure 1 | Resonator structure and transverse manifold geometry.

a, Top, ray trajectories (black lines) in a curved mirror resonator oscillate transversely (green arrows). In a particular transverse plane, the stroboscopic time evolution of the ray positions samples a harmonic oscillator trajectory (blue points). In paraxial optics, the solutions for the transverse modes are Hermite–Gauss profiles (red curve). The transverse degrees of freedom of a resonator are precisely those of a 2D quantum harmonic oscillator (bottom). **b**, Top, as a four mirror resonator is made non-planar (purple arrows), the light rays are induced to rotate (blue arrow) about the optic axis. In the transverse plane (represented below), this corresponds to flattening the 2D harmonic potential (centrifugal force) and the introduction of an effective magnetic field (Coriolis force).

c, Our non-planar resonator consists of four mirrors (blue and purple) in a stretched tetrahedral configuration of on-axis length L_a and opening half-angle θ . The image rotates about the optic axis (red) on every round trip. **d**, Left, we depict the transverse plane at the resonator waist pierced by a uniform perpendicular (along \hat{z}) magnetic field \mathbf{B} of magnitude B , and show a generic profile (red curve) with three-fold symmetry. When the plane is cut arbitrarily into three equal sections, the entire profile is fully determined within any one-third section of the plane: when a trajectory leaves one side of a section, it reappears on the other side. Each section may be wrapped into a cone on which the original profile appears once (right; this would be true for any discrete rotational symmetry). The effective magnetic field is everywhere perpendicular to the cone's surface.

In practice, we tune our resonator to degeneracy by varying its length, which primarily changes the harmonic trapping without changing the effective magnetic field, and we track the energy spectrum and spatial profiles of resonator modes by observing the transmission of circularly polarized light with a holographically programmed spatial profile (Fig. 2, see Supplementary Information). Figure 2a shows the evolution of a number of mode energies in numerous Landau levels as we adjust the resonator length over almost a centimetre. Using the observed mode-profiles (shown as insets), we identify the four lowest modes in the figure as those comprising the lowest conical Landau level, and centre the graph on their approximate degeneracy point. Figure 2b shows high-resolution spectroscopy of a larger number of modes in the lowest Landau level near the length where the harmonic confinement is cancelled. We precisely extract the change in resonator length from the spectroscopically measured free spectral range and compensate the residual harmonic trapping to zero. At this point, the residual non-degeneracy comes from local disorder, which causes an observed level repulsion for high angular momentum states (Fig. 2b, main panel) that is not observed at lower angular momentum (Fig. 2b, top inset) as well as a significant reduction in mode lifetime (Fig. 2c). Away from degeneracy the modes are nearly ideal rings with $2\pi \times l$ phase winding (experimentally determined by varying the phase profile of the injected light, see Supplementary Information); at degeneracy these modes mix due to local disorder potentials (Fig. 2d). This effect is apparent because of the long particle lifetime (high finesse of our resonator) and, in only causing mode distortion, is qualitatively different from global potentials such as astigmatism that cause mode deconfinement (see Supplementary Information). The local disorder merely creates chiral, localized states; it does not break topological

protection so long as it only mixes modes within a single Landau level and, in an interacting system, is weaker than the interactions. This insensitivity to weak disorder is a notable advantage of our set-up as compared to, for example, injecting angular momentum modes into a two mirror resonator¹⁹ (see Supplementary Information).

To demonstrate our system's stability out to large displacements from the cone tip, Fig. 3a, b shows large-angular-momentum orbits. Figure 3a presents a large displaced state composed of modes with angular momentum up to $l \approx 60$, which exhibits three-fold symmetry and interferes with itself, producing a strongly fringed pattern due to the rapid phase winding of each ring. Figure 3b unwraps another large-angular-momentum mode showing that if an orbit encircles the cone tip, then it must do so three times, as a consequence of the three-fold symmetry.

Remarkably, photons in our resonator may live on three distinct cones, distinguished by additional magnetic flux threaded through their tips. To understand this, note that the planar lowest Landau level may be spanned by angular momentum states $\psi_l(z = \frac{x+iy}{w_0}) \propto z^l \exp(-|z|^2)$ for $l=0, 1, 2, \dots$, with the transverse position vector $\mathbf{r} = (x, y)^T$. In our resonator these are partitioned into three separately degenerate sets corresponding to lowest Landau levels on different cones. These sets are the $l=0, 3, 6, \dots$ modes, the $l=1, 4, 7, \dots$ modes, and the $l=2, 5, 8, \dots$ modes and satisfy the angular symmetry condition $\psi_l(\theta + 2\pi/3) = e^{2\pi i c/3} \psi_l(\theta)$, where $c=0, 1$, or 2 is the lowest angular momentum state in the set and serves as the cone's label. $c=0$ defines the symmetry relation that describes an unthreaded cone; with $c \neq 0$, the cone has an additional Aharonov–Bohm phase arising from $c/3$ magnetic flux quanta threaded through its tip (Fig. 3c). Angular

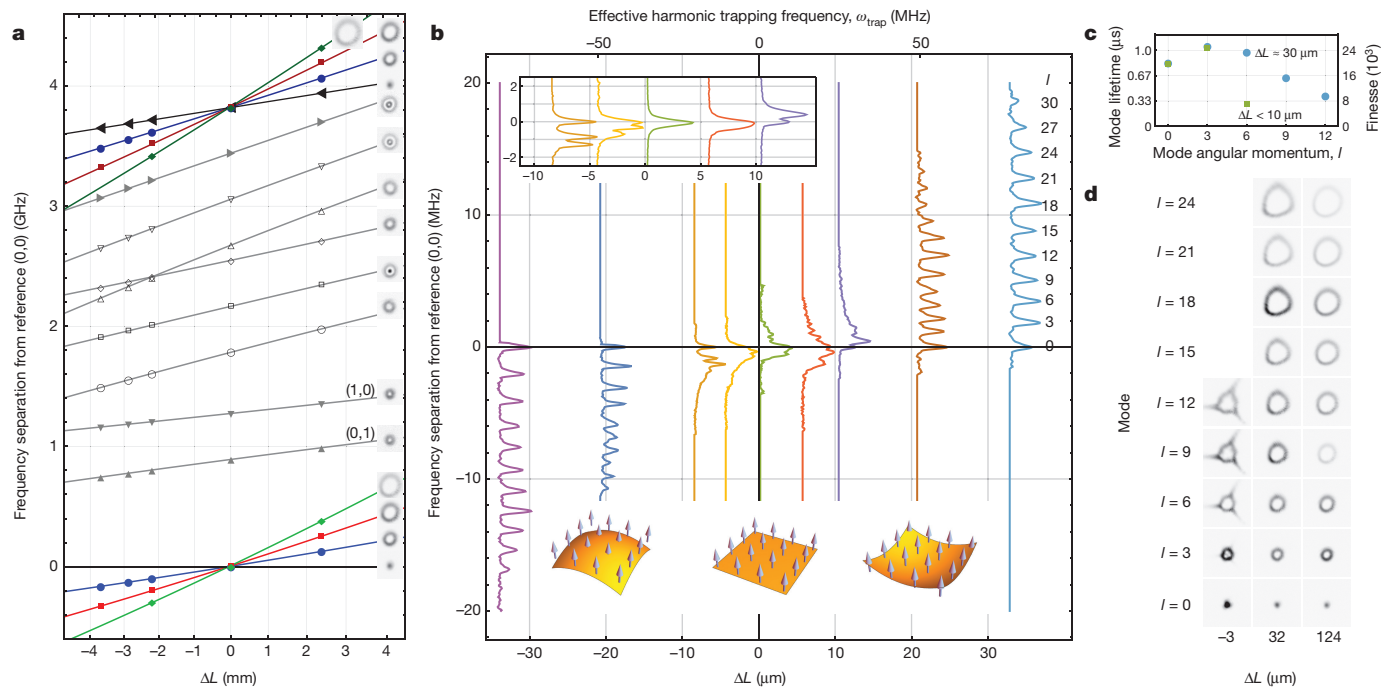


Figure 2 | Building a Landau level. The modes of our resonator follow the Fock–Darwin Hamiltonian of a massive, harmonically trapped particle in magnetic field: the magnetic field creates a ladder of Landau levels uniformly spaced by the cyclotron frequency, ω_c , while the harmonic trap of frequency ω_{trap} uniformly splits levels within each Landau level by $\omega_{\text{trap}}/\omega_c$ (see Supplementary Information). We probe this spectrum versus resonator length L_{rt} and demonstrate that, for each L_{rt} , the spectrum is determined by two energies $\nu_{(1,0)}$ and $\nu_{(0,1)}$ according to $\nu_{(\alpha,\beta)} = \alpha\nu_{(1,0)} + \beta\nu_{(0,1)} \bmod \nu_{\text{FSR}}$, where $\omega_c = 2\pi \times \nu_{(1,1)}$ gives the cyclotron frequency and $\omega_{\text{trap}}/\omega_c = 2\pi \times \nu_{(3,0)}$ provides the harmonic trapping frequency. Furthermore, fine-tuning L_{rt} drives ω_{trap} to zero, bringing specific sets of angular momentum eigenmodes into degeneracy, thereby forming Landau levels. **a**, The frequency separations between several modes and a reference $l=0$ mode are plotted as the harmonic confinement is coarsely tuned relative to an approximately degenerate reference length $L_{\text{rt}} = 78.460$ mm (corresponding free spectral range $\nu_{\text{FSR}} = 3.8209$ GHz). Solid lines are obtained as integer linear combinations of fits to the modes labelled (1,0) and (0,1) and the free spectral range. For details on mode indexing, see Supplementary Information. **b**, Main panel, we plot the transmission spectrum of the first ~10 modes in the lowest Landau level against small deviations from nominal degeneracy. Top inset, low order

modes become degenerate to within a resonator linewidth, $\kappa \approx 200$ kHz, while in the main panel, we observe weak level repulsion (approximately equal to the resonator linewidth) in the higher order modes consistent with mode mixing due to mirror imperfections of $\sim \lambda/5,000$. ω_{trap} is presented on the upper horizontal axis. Bottom insets, as the resonator is tuned through degeneracy, the harmonic potential (orange surface) changes sign, while the magnetic field (blue arrows) remains nearly unchanged. **c**, The lifetimes (and corresponding finesse) of representative modes decrease for higher mode numbers both away from degeneracy (blue circles) and near degeneracy (green squares). Here ΔL is the offset of the round-trip resonator length from nominal degeneracy. **d**, With significant residual harmonic trapping ($\Delta L = 124 \mu\text{m}$), angular momentum modes are simple rings. As the trapping is reduced ($\Delta L = 32 \mu\text{m}$), high angular momentum modes begin to mix owing to local disorder. When the trapping is precisely cancelled ($\Delta L = -3 \mu\text{m}$), mirror imperfection consistent with a single nanoscopic scratch dramatically alters the modes' shape away from the predicted near-Laguerre–Gauss profiles. Even the first resonator mode is noticeably triangular, indicating at least a mixing of Laguerre–Gauss $l=0$ and $l=3$ modes. Overcoming this disorder necessitates only $\sim \text{MHz}$ photon–photon interactions to explore strongly correlated physics.

momentum states encircling the cone tip enclose this flux three times, so states experience integer flux, reflected in their \sqrt{l} radial extension.

Away from the apex, photons on each cone behave as in a flat space lowest Landau level. In Fig. 3d, we identify each cone by the lowest angular momentum state supported around its apex. Then, on each cone, we show that we can create arbitrary angular momentum states ($l=0, 1$) about displaced points so long as the displaced mode does not self intersect or encircle the cone tip. Beyond reflecting the invariance of our system under magnetic translations, this permits the creation of canonical fractional quantum Hall states in a future interacting system, in addition to novel Laughlin states accessible at the cone tip (see Supplementary Information). As a visualization, Fig. 3e, f projects these displaced $l=0$ and $l=1$ modes onto a cone, further demonstrating that, away from the apex, modes on the cone closely resemble modes on a regular plane.

The topological numbers that characterize quantum Hall phases are predicted to specify the response of the photonic local density of states (LDOS) to magnetic field and spatial curvature, as described by the Wen–Zee theory^{12–15} (see Supplementary Information). We perform an experimental test of this theory by measuring the LDOS (Fig. 3g–i)

via transmission images of each state in the relevant weakly split Landau level and summing these images (see Supplementary Information). We then compare the LDOS near the cone tip with the flat space density away from the tip (within each panel Fig. 3g–i) and compare the LDOS with different quantities of flux threaded (between panels Fig. 3g–i). We clearly observe a density build-up for the $c=0$ cone; however, we find a vanishing LDOS on the other two cones, reflecting additional magnetic flux threaded through their tips equal to $-\Phi_0/3$ and $-2\Phi_0/3$, where Φ_0 is the magnetic flux quantum (Fig. 3c). According to the Wen–Zee theory, the expected excess state number is given by $\delta N = \frac{2}{3}\bar{s} - \frac{c}{3}$, where $c/3$ is the number of flux quanta threaded through the cone tip and \bar{s} is a parameter called the mean orbital spin that characterizes particles' coupling to spatial curvature and is predicted to be $1/2$ for the lowest Landau level¹² (see Supplementary Information). We therefore expect $\delta N = 1/3, 0$, and $-1/3$ of a state near the tips of the $c=0, 1$, and 2 cones, respectively. By integrating the measured LDOS excess or deficit near the apex, we measure the state number excess to be $0.31(2)$ on the $c=0$ cone, $-0.02(1)$ on the $c=1$ cone, and $-0.35(2)$ on the $c=2$ cone, yielding the experimentally measured value $\bar{s} = 0.47(1)$. We find quantitative agreement between our measured results and the Wen–Zee theory.

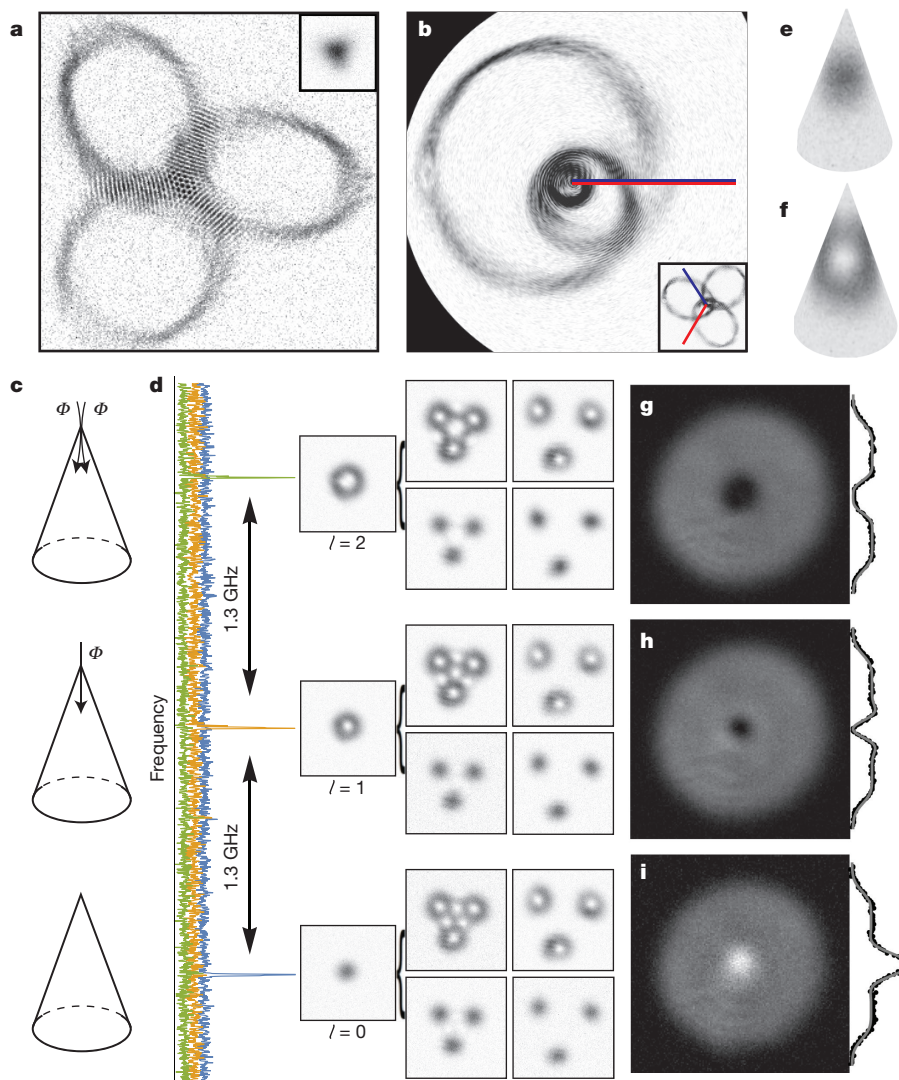


Figure 3 | Photonic lowest Landau levels on a cone. **a**, At degeneracy, all resonator modes display three-fold symmetry. We present a very large displaced angular momentum mode with radial extent up to 8 times the mode waist, w_0 , implying that ~ 20 modes must be degenerate. The rapid phase winding for large l modes causes the strong fringing pattern when the mode self-interferes. Inset, an $l=0$ mode at the same scale. **b**, We project another large-angular-momentum mode onto a cone and view it from above the apex. We observe a general property that circular orbits must encircle the cone apex either zero or three times. Inset, the original image of the mode. The pair of rays overlaying the inset image corresponds to the cut in the main image. **c**, The twisted resonator corresponds to Landau levels on three cones with differing quantities of magnetic flux threaded through the tip. The cone built out of $l=0, 3, 6, \dots$ has no flux threading; the cone built out of $l=1, 4, 7, \dots$ is threaded by $\Phi_0/3$; and the cone built out of $l=2, 5, 8, \dots$ is threaded by $2\Phi_0/3$, where Φ_0 is the magnetic flux quantum. **d**, With the resonator tuned to degeneracy, we identify the energies of the $l=c$ modes for $c=0, 1$ or 2 by the transmission peaks (blue, orange, and green curves, respectively) that correspond to the correct observed transmitted modes' profiles (single images, labelled). The degenerate sets starting with these modes each form a lowest Landau level on different cones. Except at the apex, each cone is flat, so away from

the tip each lowest Landau level supports modes of—and therefore the dynamics of—a planar lowest Landau level with $l=0, 1, 2, \dots$ defined about a displaced point. On each cone, we show displaced $l=0$ (bottom two) and $l=1$ (top two) modes. For large displacements (right two), these modes are undistorted; however, for small displacements (left two), where there is significant mode amplitude at the tip, we observe distortions due to self-interference, similar to **a**, **e**, **f**. Displaced $l=0$ and $l=1$ modes from **d** are projected onto a cone to show how observed mode images may be interpreted on a conical surface. **g–i**, We explore the effects of curvature and flux threading near the tip by measuring the local density of photonic states. For the $c=0$ cone (**i**), we find an approximately threefold increase in local state density near the cone apex above a constant background plateau of density. This corresponds to an additional one-third of a state localized near the apex. For the cones with $c=1$ and 2 (**h** and **g**, respectively), we find a vanishing local density of states near the apex, reflecting the negative magnetic flux threading through the cone apex. Each unit of flux removes one-third of a state local to the apex so that the $c=1$ cone has no additional states, and the $c=2$ cone is missing one-third of one state. The data to the right display a slice through the middle of each image; the grey curves are fits to the expected analytic form (see Supplementary Information).

We have demonstrated a synthetic magnetic field for continuum photons. Furthermore, we have created an integer quantum Hall system in curved space, a long-standing challenge in condensed matter physics. We can extend our tests of the Wen–Zee theory by measuring fractional state number excess in higher Landau levels and examining the connection between the mean orbital spin and the Hall viscosity³⁰ (see Supplementary Information). Our approach clears a path to the

photonic fractional quantum Hall regime, as it is compatible with Rydberg-mediated strong photon–photon interactions¹⁶, and does not require the low particle densities (and thus weakened interactions) necessary to map Laughlin physics onto a lattice. Simply avoiding the cone apex will allow the spectroscopic creation and detection of flat space fractional quantum Hall states such as the Laughlin wavefunction (see Supplementary Information), while exploring the apex will afford

the opportunity to investigate the interplay of geometry and topology in strongly correlated quantum materials.

Received 24 November 2015; accepted 9 March 2016.

Published online 8 June 2016.

1. Peyronel, T. *et al.* Quantum nonlinear optics with single photons enabled by strongly interacting atoms. *Nature* **488**, 57–60 (2012).
2. Gopalakrishnan, S., Lev, B. L. & Goldbart, P. M. Emergent crystallinity and frustration with Bose–Einstein condensates in multimode cavities. *Nat. Phys.* **5**, 845–850 (2009).
3. Baumann, K., Guerlin, C., Brennecke, F. & Esslinger, T. Dicke quantum phase transition with a superfluid gas in an optical cavity. *Nature* **464**, 1301–1306 (2010).
4. Cooper, N. R. & Dalibard, J. Reaching fractional quantum Hall states with optical flux lattices. *Phys. Rev. Lett.* **110**, 185301 (2013).
5. Carusotto, I. & Ciuti, C. Quantum fluids of light. *Rev. Mod. Phys.* **85**, 299–366 (2013).
6. Jia, N., Owens, C., Sommer, A., Schuster, D. & Simon, J. Time- and site-resolved dynamics in a topological circuit. *Phys. Rev. X* **5**, 021031 (2015).
7. Otterbach, J., Ruseckas, J., Unanyan, R. G., Juzeliūnas, G. & Fleischhauer, M. Effective magnetic fields for stationary light. *Phys. Rev. Lett.* **104**, 033903 (2010).
8. Wang, Z., Chong, Y., Joannopoulos, J. D. & Soljačić, M. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009).
9. Yuan, J. *et al.* Nonplanar ring resonator modes: generalized Gaussian beams. *Appl. Opt.* **46**, 2980–2989 (2007).
10. Cooper, N. R. Rapidly rotating atomic gases. *Adv. Phys.* **57**, 539–616 (2008).
11. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
12. Wen, X. G. & Zee, A. Shift and spin vector: new topological quantum numbers for the Hall fluids. *Phys. Rev. Lett.* **69**, 953–956 (1992).
13. Hoyos, C. & Son, D. T. Hall viscosity and electromagnetic response. *Phys. Rev. Lett.* **108**, 066805 (2012).
14. Abanov, A. G. & Gromov, A. Electromagnetic and gravitational responses of two-dimensional noninteracting electrons in a background magnetic field. *Phys. Rev. B* **90**, 014435 (2014).
15. Can, T., Laskin, M. & Wiegmann, P. Fractional quantum Hall effect in a curved space: gravitational anomaly and electromagnetic response. *Phys. Rev. Lett.* **113**, 046803 (2014).
16. Sommer, A., Büchler, H. P. & Simon, J. Quantum crystals and Laughlin droplets of cavity Rydberg polaritons. Preprint at <http://arXiv.org/abs/1506.00341> (2015).
17. Umucalilar, R. O., Wouters, M. & Carusotto, I. Probing few-particle Laughlin states of photons via correlation measurements. *Phys. Rev. A* **89**, 023803 (2014).
18. Paredes, B., Fedichev, P., Cirac, J. I. & Zoller, P. 1/2-Anyons in small atomic Bose–Einstein condensates. *Phys. Rev. Lett.* **87**, 010402 (2001).
19. Umucalilar, R. O. & Carusotto, I. Many-body braiding phases in a rotating strongly correlated photon gas. *Phys. Lett. A* **377**, 2074–2078 (2013).
20. Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
21. Wang, Z., Chong, Y. D., Joannopoulos, J. D. & Soljačić, M. Reflection-free one-way edge modes in a gyromagnetic photonic crystal. *Phys. Rev. Lett.* **100**, (2008).
22. Rechtsman, M. C. *et al.* Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013).
23. Hafezi, M., Mittal, S., Fan, J., Migdall, A. & Taylor, J. M. Imaging topological edge states in silicon photonics. *Nat. Photon.* **7**, 1001–1005 (2013).
24. Rechtsman, M. C. *et al.* Strain-induced pseudomagnetic field and photonic Landau levels in dielectric structures. *Nat. Photon.* **7**, 153–158 (2013).
25. Karzig, T., Bardyn, C.-E., Lindner, N. H. & Refael, G. Topological polaritons. *Phys. Rev. X* **5**, 031001 (2015).
26. Longhi, S. Synthetic gauge fields for light beams in optical resonators. *Opt. Lett.* **40**, 2941–2944 (2015).
27. Klaers, J., Schmitt, J., Vewinger, F. & Weitz, M. Bose–Einstein condensation of photons in an optical microcavity. *Nature* **468**, 545–548 (2010).
28. Sommer, A. & Simon, J. Engineering photonic Floquet Hamiltonians through Fabry–Pérot resonators. *New J. Phys.* **18**, 035008 (2016).
29. Schweikhard, V., Coddington, I., Engels, P., Mogendorff, V. P. & Cornell, E. A. Rapidly rotating Bose–Einstein condensates in and near the lowest Landau level. *Phys. Rev. Lett.* **92**, 040404 (2004).
30. Read, N. Non-Abelian adiabatic statistics and Hall viscosity in quantum Hall states and $p_x + ip_y$ paired superfluids. *Phys. Rev. B* **79**, 045308 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge conversations with I. Carusotto, M. Levin and P. Wiegmann. This work was supported by DOE, DARPA and AFOSR. A.G. acknowledges the support of the Kadanoff Center for Theoretical Physics. A.R. acknowledges support from ARO through an NDSEG fellowship.

Author Contributions The experiment was designed and built by N.S., J.S., A.R. and A.S. Measurement and analysis of the data was performed by N.S. Theoretical development and interpretation of results were performed by J.S., A.S., N.S. and A.G. All authors contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S. (simonjon@uchicago.edu).

Switching stiction and adhesion of a liquid on a solid

Stijn F. L. Mertens^{1,2*}, Adrian Hemmi^{3*}, Stefan Muff³, Oliver Gröning⁴, Steven De Feyter¹, Jürg Osterwalder³ & Thomas Greber³

When a gecko moves on a ceiling it makes use of adhesion and stiction. Stiction—static friction—is experienced on microscopic and macroscopic scales and is related to adhesion and sliding friction¹. Although important for most locomotive processes, the concepts of adhesion, stiction and sliding friction are often only empirically correlated. A more detailed understanding of these concepts will, for example, help to improve the design of increasingly smaller devices such as micro- and nanoelectromechanical switches². Here we show how stiction and adhesion are related for a liquid drop on a hexagonal boron nitride monolayer on rhodium³, by measuring dynamic contact angles in two distinct states of the solid–liquid interface: a corrugated state in the absence of hydrogen intercalation and an intercalation-induced flat state. Stiction and adhesion can be reversibly switched by applying different electrochemical potentials to the sample, causing atomic hydrogen to be intercalated or not. We ascribe the change in adhesion to a change in lateral electric field of in-plane two-nanometre dipole rings⁴, because it cannot be explained by the change in surface roughness known from the Wenzel model⁵. Although the change in adhesion can be calculated for the system we study⁶, it is not yet possible to determine the stiction at such a solid–liquid interface using *ab initio* methods. The inorganic hybrid of hexagonal boron nitride and rhodium is very stable and represents a new class of switchable surfaces with the potential for application in the study of adhesion, friction and lubrication.

Every object at rest sticks with some adhesion to its substrate. If this object is moved, then a force must act on it. This force has to overcome the stiction threshold, above which the object starts to move and below which it sticks to the substrate. This fundamentally simple principle indicates a relationship between adhesion and stiction that is valid for a range of length scales from single-atom manipulation⁷ up to the everyday experience of moving ourselves. Stiction is also related to sliding friction^{1,8} and surface diffusion⁹. Empirically, these properties are connected via dimensionless coefficients that relate, for example, the diffusion barrier to the adsorption energy or the sliding friction force to the load. The advent of single-atom probes on surfaces and the measurement of the heat of adsorption have made the Evans–Polanyi relation, which postulates proportionality between the activation energy for diffusion and the adsorption energy, experimentally accessible¹⁰. For macroscopic objects, on the other hand, Amontons' first law, which states that sliding friction is proportional to the load, is much more complex because of the many length scales involved. This complexity also applies to stiction and, consequently, an *ab initio* understanding is difficult¹¹. For such problems we rely on clear-cut responsive-surface model systems to study the influence of microscopic effects on their macroscopic expressions and vice versa.

For responsive surfaces, a collective change of a microscopic atomic or molecular parameter triggers a macroscopic property change such as the wetting angle of a liquid; such surfaces are central to nanoscience and smart-materials research^{12–14}. Typically, the responsiveness is invoked by a change of conformation or charge state of organic

molecules at the solid–liquid interface, which can be triggered by light¹², temperature¹³ or electric fields¹⁴. The organic molecules at the basis of the switching, however, render these surfaces quite fragile and limited to near-ambient temperatures. By contrast, the inorganic hexagonal boron nitride/rhodium (h-BN/Rh) hybrid we focus on has very high thermal and chemical stability (see Methods), and therefore signifies a class of responsive surfaces with considerable technological promise. The stability of the h-BN/Rh system is related to the covalently bonded network of the h-BN 'skin' in comparison with the weaker supramolecular interactions in self-assembled monolayers.

The boron nitride nanomesh is a corrugated monolayer of h-BN on rhodium^{3,15}: 13×13 h-BN units form a superhoneycomb structure on 12×12 Rh(111) unit cells, in which particularly strong lateral electric fields exist that are, for instance, decisive for the self-assembly of molecules⁴. The nanomesh structure is stable in vacuum up to 1,000 °C and survives immersion into liquids^{4,16}. In ultrahigh vacuum, it was found that the boron nitride layer could be made flat by intercalation of atomic hydrogen¹⁷. Here, we demonstrate this effect under electrochemical conditions and show that h-BN/Rh(111) is a surface with switchable wetting and adhesion.

The electrochemical switching of the surface texture is based on intercalation of atomic hydrogen, as is demonstrated in Fig. 1. Figure 1a shows cyclic voltammograms for a clean Rh(111) film and a h-BN nanomesh sample in perchloric acid solution in a sessile drop configuration¹⁸. In the negative scan direction, the Rh(111) reference voltammogram exhibits the typical atomic hydrogen adsorption peak just before the onset of molecular hydrogen evolution¹⁹. On h-BN/Rh(111), this peak occurs at less negative potentials, which suggests that the process is energetically slightly more favourable on the nanomesh. During the reverse scan, the single desorption peak on Rh(111) shows a double peak for h-BN/Rh(111), which indicates a two-stage desorption process. Integration of the hydrogen adsorption peak yields $75 \mu\text{C cm}^{-2}$, which amounts to one-third of the bare Rh(111) measurement where one monolayer of hydrogen may be adsorbed, and confirms earlier observations^{16,17}. On continued cycling (Fig. 1b), a slight sharpening of the adsorption peak occurs; the charge related to the peak integral changes by less than 15%, demonstrating the stability of the h-BN overlayer upon cycling.

The changes in nanotexture that accompany the hydrogen adsorption and desorption were visualized directly by *in situ* electrochemical scanning tunnelling microscopy (EC-STM). Figure 1c–f shows the same area for different substrate potentials. Initially, at potential E_1 (Fig. 1c), the hexagonal pattern of pores separated by wires is observed^{16,17}. The 3.2-nm periodicity can also be seen in the autocorrelation of a given section from one terrace, and the corrugation is reflected by the root-mean-square (r.m.s.) roughness. Crossing of the hydrogen adsorption peak (potential E_2 , Fig. 1d) leads to flattening of the surface, which is reflected in the vanishing of the superstructure and a decrease in the r.m.s. roughness. A tip change between Fig. 1c and Fig. 1e caused a contrast inversion of the pore, which, however, does not affect the r.m.s. analysis. When the potential is set back to

¹Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium. ²Institut für Angewandte Physik, Technische Universität Wien, Wiedner Hauptstrasse 8–10/E134, 1040 Wien, Austria. ³Physik-Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. ⁴Empa, Swiss Federal Laboratories for Materials Science and Technology, Überlandstrasse 129, 8600 Dübendorf, Switzerland.

*These authors contributed equally to this work.

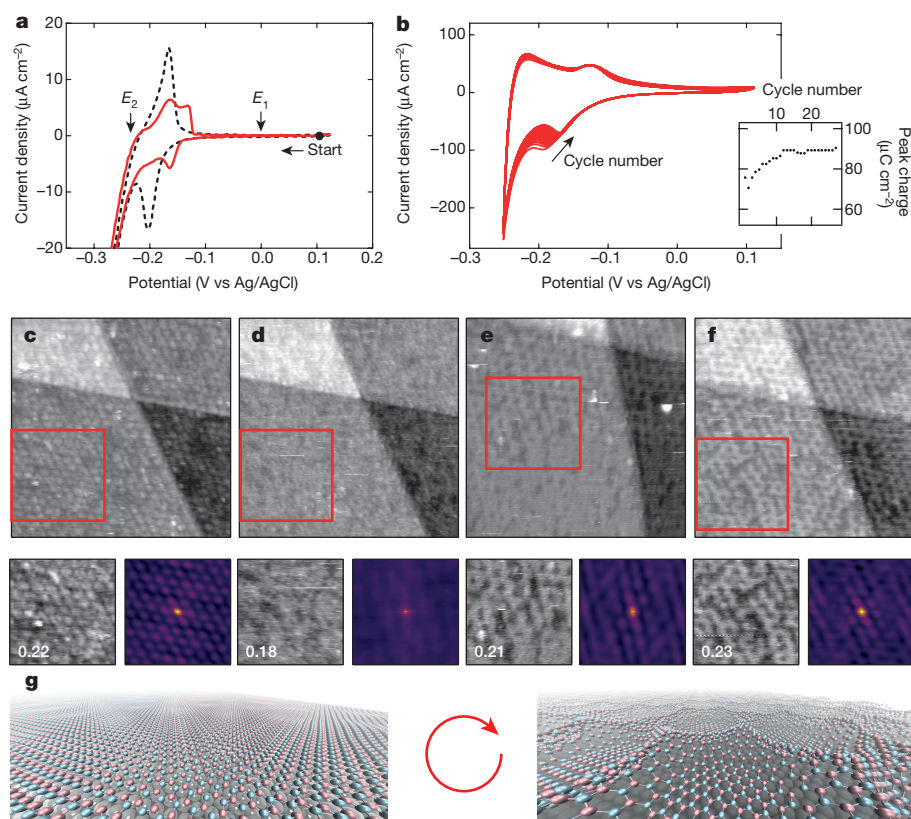


Figure 1 | Voltammetry and electrochemical scanning tunnelling microscopy. **a**, Cyclic voltammogram of h-BN nanomesh on Rh(111) (red trace) and of Rh(111) (black dashed trace) in contact with a sessile drop of 0.1 M HClO₄. Scan rate, 1 mV s⁻¹. The hydrogen adsorption peak is found at -150 mV for h-BN/Rh(111) and at -200 mV for Rh(111). $E_1 = 0$ V and $E_2 = -0.25$ V are the substrate potentials at which *in situ* STM was performed. **b**, Stability of the hydrogen adsorption/desorption peaks over multiple cycles (28 cycles shown, recorded in a standard electrochemical cell; scan rate, 10 mV s⁻¹). The charge of the hydrogen adsorption peak as a function of cycle number is shown in the inset. **c–f**, Sequence of *in situ* STM images for a region with three atomic terrace levels, at various substrate potentials: $E_1 = 0$ V (**c**); $E_2 = -0.25$ V (**d**); switching $E_2 \rightarrow E_1$ (**e**; image scanned from bottom to top); and $E_1 = 0$ V after recovery of corrugation (**f**). The red square highlights the same area of the sample in all panels; close-ups ($1.2 \times$ magnification) are shown in the bottom-left panels. The white numbers are the r.m.s. roughness in units of the terrace height. The bottom-right panels show the autocorrelations of the bottom-left panels and reveal, if present, the 3.2-nm periodicity of the h-BN superstructure. Image sizes, 66 nm \times 66 nm; tunnelling current, 0.1 nA; tip potential fixed at -0.45 V. **g**, Three-dimensional representation of the flat (left) and the corrugated h-BN layer (right). N, sky blue; B, pink; Rh, dark grey; H, white. For clarity, the heights of N and B above the Rh top layer have been stretched by a factor of three. Coordinates taken from refs 17 and 29.

the initial value, the corrugation gradually reappears (Fig. 1e), which indicates that deintercalation is slower than the forward process (see Methods and Extended Data Fig. 2). After the cycle (Fig. 1f), the superhoneycomb shows different imperfections (for example, some pores appear connected). The physical picture that emerges is illustrated in Fig. 1g: electrochemical intercalation of hydrogen reversibly flattens the h-BN/Rh nanomesh, and in doing so switches off the dipole rings in the surface.

To prove unequivocally that hydrogen intercalation is at the basis of the observed switching effect, we conducted an ambient-to-vacuum transfer experiment with the goal of quantifying the intercalated hydrogen. The electrochemical treatment was performed in heavy water, and the h-BN/Rh(111) sample was extracted from the solution under potential control and transferred in less than 10 min to vacuum (see Methods) where thermal desorption spectroscopy (TDS) was carried out, as shown in Fig. 2. Deuterium desorption from the sample held at potential E_2 (flat state) is obvious, whereas no substantial desorption took place from a sample that was kept at E_1 (corrugated) throughout. The desorption temperature indicates slightly higher binding energy as compared to hydrogen on bare Rh(111) (ref. 20). This further confirms hydrogen intercalation, because hydrogen (or deuterium) on top of h-BN would not bind as strongly as on rhodium and would not survive the transfer from the liquid to the vacuum. Our experiments do not show atomistic details of the way in which the hydrogen intercalates, although it has been claimed²¹ that protons can pass through an intact h-BN layer, and that this process is facilitated by the presence of a platinum group metal (see Methods). In this scenario, the hydrated protons in the liquid electrolyte are probably stripped of their solvation shell the instant intercalation occurs, or a Grotthuss-type transport mechanism may be involved²².

Stiction of the electrolyte on the h-BN surface is studied by measuring dynamic contact angles, which differ from the equilibrium value

in Young's equation. This difference can be directly observed if a liquid drop on a solid loses volume by evaporation, where the drop footprint will start to move only below a critical contact angle (the receding angle). Likewise, when the volume of the drop is gradually increased, the footprint moves when the advancing contact angle is reached^{23,24}. The difference in advancing and receding angles and the concomitant contact angle hysteresis are a macroscopic expression of adhesion and stiction. In Fig. 3, we show dynamic contact angle measurements for

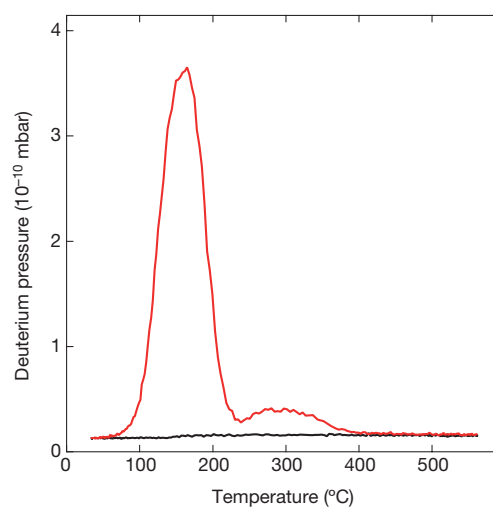


Figure 2 | Deuterium thermal desorption spectra. h-BN/Rh(111) was exposed to 0.1 M DClO₄ in D₂O at substrate potentials of $E_1 = 0$ V (black trace) and $E_2 = -0.25$ V (red trace). After loading the samples in the electrolyte they were transferred to ultrahigh vacuum for TDS (see Methods). Heating rate, 0.9 K s⁻¹.

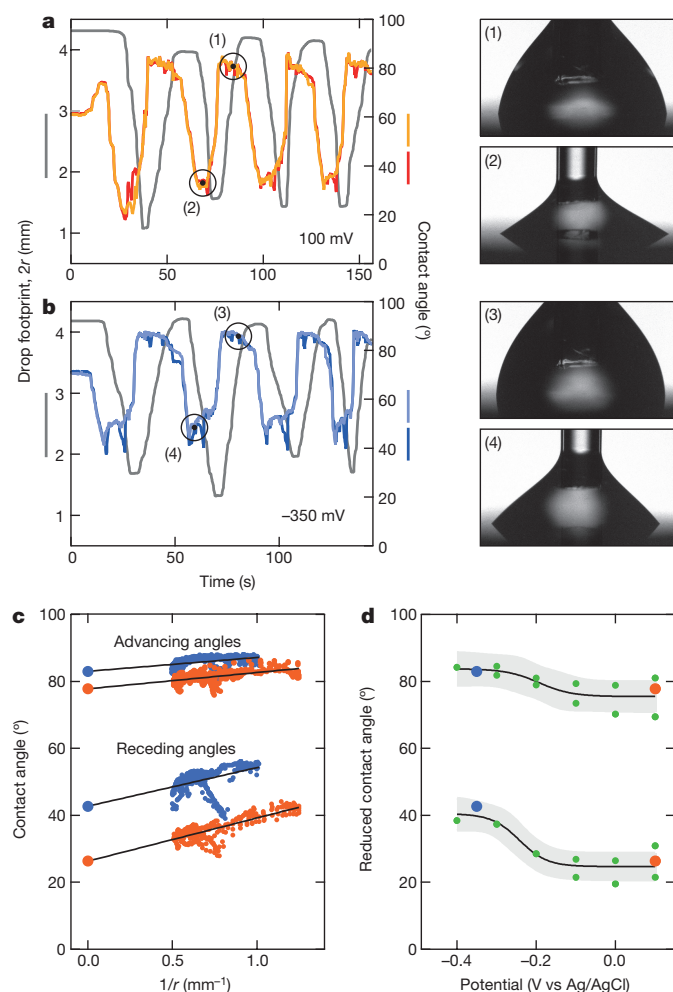


Figure 3 | Dynamic contact angle measurements. **a**, Left (orange) and right (red) contact angles (right axis) and drop footprints (with diameters of $2r$; grey, left axis) as functions of time. Pictures of an advancing (1) and a receding (2) drop in the transients at 100 mV (corrugated state; indicated by the circled points in the left panel) are shown on the right. **b**, Left (dark blue) and right (light blue) contact angles and drop footprints as functions of time. Pictures of an advancing (3) and receding (4) drop in the transients at -350 mV (flat state; indicated by the circled points in the left panel) are shown on the right. **c**, Contact angles in **a** (100 mV, red) and **b** (-350 mV, blue) versus $1/r$ for times above 40 s. Extrapolation to $1/r=0$ (infinite drop footprint) yields reduced contact angles that are independent of the contact line. **d**, Reduced receding (bottom) and advancing (top) angles for different applied potentials, where the hydrogen intercalation peak is found at about -0.2 V. The large red and blue points correspond to the values determined in **c**. The solid lines are sigmoid fits through all experimental points (green circles), from which values of θ_a and θ_r for the corrugated ($76.0^\circ \pm 1.2^\circ$ and $24.5^\circ \pm 1.4^\circ$, respectively) and the flat ($84.3^\circ \pm 2.3^\circ$ and $40.1^\circ \pm 3.2^\circ$, respectively) state are determined. The grey band indicates the 1-s.d. prediction interval.

applied potentials around the hydrogen adsorption peak. The angles as determined from inflating and deflating the drops in the corresponding videos (Supplementary Videos 1, 2) are shown for a potential above (Fig. 3a) and below (Fig. 3b) the hydrogen peak, that is, for the corrugated and the flat surface, respectively. The drop volume was changed periodically and distinct receding angles θ_r and advancing angles θ_a were observed. Four positions are shown as photographs of the electrolyte drop on a h-BN/Rh(111) sample with the capillary, which is 0.85 mm in diameter, entering the drop from the top. The experiments show distinct hysteresis in the contact angles of about 60°. The observed angles also depend on residual defects, which is reflected in the asymmetry between the left and the right footprint, and in sudden angle

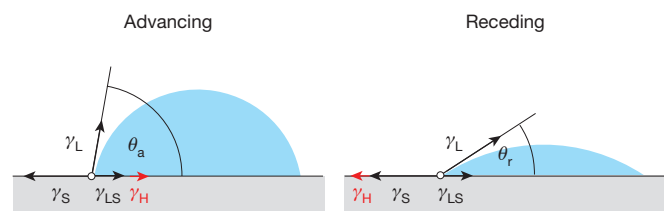


Figure 4 | Wetting angle hysteresis and stiction. The footprint of a drop starts to move with an advancing (left; θ_a) or a receding (right; θ_r) angle, if the friction-induced lateral tension exceeds the stiction threshold γ_H . The equilibrium angle from Young's equation θ_0 lies between θ_a and θ_r . γ_s , γ_{LS} and γ_L are the corresponding solid-gas, liquid-solid and liquid-gas interface energies, respectively.

changes that are reproduced for subsequent cycles. All the effects due to residual imperfections, however, are distinct from the effect of the electrochemical potential. The size dependence of the wetting hysteresis is a consequence of the line tension, that is, the work required to change the length of the drop perimeter²⁵. If the wetting angles are plotted against the inverse of the footprint radius (Fig. 3c), then the line tension and the extrapolation of the wetting angle for infinite footprints may be inferred. In Fig. 3d, these extrapolated advancing and receding angles are plotted for a large set of experiments as a function of the applied potential, correlating the applied substrate potential with the measured contact angles. Clearly, the hystereses for potentials above -200 mV are different from those below this threshold of the hydrogen adsorption peak. Further experiments indicate that neither visible evolution of molecular hydrogen (where the reduction current increases exponentially in the voltammogram) nor electrowetting effects^{26,27} can be responsible for the observed changes in contact angle.

The concept for stiction and the concomitant contact angle hysteresis is sketched in Fig. 4: the static friction σ balances the forces that distort the Young equilibrium $\gamma_s - \gamma_{LS} - \gamma_L \cos(\theta_0) = 0$ (in which γ_s , γ_{LS} and γ_L are the interfacial energies or tensions of the solid-gas, liquid-solid and gas-liquid interfaces, and θ_0 is the Young equilibrium angle). The static friction σ acts parallel to the surface, as do γ_s and γ_{LS} , and invokes a new equilibrium angle θ_σ . The new equilibrium angle is determined by

$$\gamma_s - \gamma_{LS} - \gamma_L \cos(\theta_\sigma) + \sigma = 0$$

The static friction σ may, however, not exceed a critical value $\pm \gamma_H$, above which the footprint of the drop starts to move with an advancing or a receding angle. In Fig. 3c, d, the size-independent receding and advancing angles (that is, the extrapolation to infinite footprints) θ_r and θ_a are indicated for the corrugated and the flat h-BN layer. Their values are given by

$$\cos(\theta_{r,a}) = \cos(\theta_0) \mp \frac{\gamma_H}{\gamma_L}$$

in which $\gamma_H \geq \sigma$ is the stiction threshold. With the knowledge of γ_L (72 mJ m^{-2} for water), which does not depend on the surface of the solid, and the determination of θ_a and θ_r , γ_H and θ_0 may be directly determined.

Our experiments indicate that the hydrogen-intercalation-induced flattening of the h-BN nanomesh increases θ_0 from $54.9^\circ \pm 0.8^\circ$ to $64.5^\circ \pm 1.7^\circ$. Using the Young-Dupré equation we determine the work of adhesion per unit area w_{LS} from γ_L and the Young equilibrium angle θ_0 (ref. 28):

$$w_{LS} = \gamma_L + \gamma_s - \gamma_{LS} = \gamma_L [1 + \cos(\theta_0)]$$

The work of adhesion decreases from $(1.58 \pm 0.01)\gamma_L$ by $9.2\% \pm 1.8\%$ in going from the corrugated to the flat layer. The stiction threshold γ_H of $(0.33 \pm 0.01)\gamma_L$ and $(0.33 \pm 0.03)\gamma_L$ for corrugated and flat layers,

respectively, is (within the error bars) the same for the two interfaces. The change in work of adhesion is incompatible with a Wenzel model⁵, according to which a change in effective surface area is responsible for a change in wetting angle: the effective surface area of the corrugated h-BN is 3% or 0.5% larger than that of the flat layer if a calculated corrugation height of 316 pm (ref. 29) or 83 pm (ref. 17) is used. Thus, to explain the magnitude of the change in adsorption energy, an electronic effect must amplify the change induced by the corrugation. We argue that this electronic effect is due to the loss of the lateral electric fields (2-nm dipole rings)⁴ in the hydrogen-intercalated surface. The ratio of adsorption energies can be tested by comparing them to water adsorption³⁰. Taking the calculated adsorption energies of water³¹ for different adsorption sites within the h-BN/Rh(111) supercell, we calculate the average adsorption energies of the corrugated and the flat state. If we take for the corrugated case the portion of the wire, pore (hole) and rim regions from Xe-adsorption⁴, and set the flat state equivalent to the unfavourable adsorption on the wires, then we find an average adsorption energy that is 18% lower for water monomers and 9% lower for hexamers on intercalated h-BN compared to the h-BN nanomesh. The comparison to the experiment is much better for hexamers, which makes sense because hexamers represent the simplest possible 'drop' in which only every second water molecule makes a hydrogen bond to the boron nitride³⁰. Beyond these estimations of the adsorption energies involved in the switching phenomenon, the comprehensively characterized structure of the nanomesh makes this surface very attractive for full-scale theoretical treatments of wetting, which is difficult for less ordered systems. Molecular dynamics results estimate a change in adhesion (adsorption energy) of 4% (ref. 6), which compares favourably with the 9.2% extracted from our wetting angle data; however, a quantitative theoretical description of stiction is still a distant prospect.

We have presented a switchable surface on which macroscopic static friction and adhesion can be linked to our understanding of water adsorption on this surface at the microscopic scale. The relatively simple structure of the h-BN nanomesh makes it amenable to accurate descriptions down to the atomic level, and an attractive model system for full-scale theoretical analyses of switchable surfaces, wetting, friction and lubrication. The fact that the intercalated surface survives emersion from the electrolyte could make this test system widely applicable for the study of friction and lubrication, even if the adhering objects differ from liquid drops. In a biological context, it might be possible to create complex multi-cellular arrangements by influencing cell migration through controlled switching between high and low adhesiveness to cells. In the context of technological advancement, the high thermal and chemical stability of the system means it could be used even in harsh environments, for example, to influence capillary action, stiction and adhesion in microfluidic or nanoelectromechanical devices. Finally, beyond the field of adhesion and friction, electrochemical hydrogen intercalation may be used to obtain freestanding boron nitride layers or, more exotically, to stably confine tritium beneath a maximally transparent single-atomic-layer moderator, as is needed to determine the neutrino mass through β -decay spectroscopy³².

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 August 2015; accepted 12 April 2016.

- Persson, B. N. J. *Sliding Friction* (Springer, 1998).
- Wagner, T. J. W. & Vella, D. Switch on, switch off: stiction in nanoelectromechanical switches. *Nanotechnology* **24**, 275501 (2013).
- Corso, M. *et al.* Boron nitride nanomesh. *Science* **303**, 217–220 (2004).
- Dil, H. *et al.* Surface trapping of atoms and molecules with dipole rings. *Science* **319**, 1824–1826 (2008).
- Wenzel, R. N. Surface roughness and contact angle. *J. Phys. Chem.* **53**, 1466–1467 (1949).
- Golze, D., Hutter, J. & Iannuzzi, M. Wetting of water on hexagonal boron nitride@Rh(111): a QM/MM model based on atomic charges derived for nanostructured substrates. *Phys. Chem. Chem. Phys.* **17**, 14307–14316 (2015).
- Ternes, M., Lutz, C. P., Hirjibehedin, C. F., Giessibl, F. J. & Heinrich, A. J. The force needed to move an atom on a surface. *Science* **319**, 1066–1069 (2008).
- Krim, J., Solina, D. H. & Chiarello, R. Nanotribology of a Kr monolayer: a quartz crystal microbalance study of atomic-scale friction. *Phys. Rev. Lett.* **66**, 181–184 (1991).
- Ehrlich, G. & Hudda, F. Atomic view of surface self-diffusion: tungsten on tungsten. *J. Chem. Phys.* **44**, 1039–1049 (1966).
- Evans, M. G. & Polanyi, M. Inertia and driving force of chemical reactions. *Trans. Faraday Soc.* **34**, 11–24 (1938).
- Zhang, Q., Qi, Y., Hector, L. G. Jr, Cagin, T. & Goddard, W. A. III. Origin of static friction and its relationship to adhesion at the atomic scale. *Phys. Rev. B* **75**, 144114 (2007).
- Ichimura, K., Oh, S.-K. & Nakagawa, M. Light-driven motion of liquids on a photoresponsive surface. *Science* **288**, 1624–1626 (2000).
- Sun, T. *et al.* Reversible switching between superhydrophilicity and superhydrophobicity. *Angew. Chem. Int. Ed.* **43**, 357–360 (2004).
- Lahann, J. *et al.* A reversibly switching surface. *Science* **299**, 371–374 (2003).
- Berner, S. *et al.* Boron nitride nanomesh: functionality from a corrugated monolayer. *Angew. Chem. Int. Ed.* **46**, 5115–5119 (2007).
- Widmer, R. *et al.* Electrolytic *in situ* STM investigation of h-BN-nanomesh. *Electrochem. Commun.* **9**, 2484–2488 (2007).
- Brugger, T. *et al.* Nanotexture switching of single-layer hexagonal boron nitride on rhodium by intercalation of hydrogen atoms. *Angew. Chem. Int. Ed.* **49**, 6120–6124 (2010).
- Willman, K. W. & Murray, R. W. Contact angle between water and a poly(vinylferrocene) film on a potential-controlled platinum electrode. *Anal. Chem.* **55**, 1139–1142 (1983).
- Sung, Y. E., Thomas, S. & Wieckowski, A. Characterization of the Rh(111) electrode by CEELS, AES, LEED, and voltammetry. Adsorption of (bi)sulfate, perchlorate, and carbon monoxide. *J. Phys. Chem.* **99**, 13513–13521 (1995).
- Colonell, J. I., Curtiss, T. J. & Sibener, S. J. Coverage dependence of the kinetics for H₂ desorption from Rh(111). *Surf. Sci.* **366**, 19–28 (1996).
- Hu, S. *et al.* Proton transport through one-atom-thick crystals. *Nature* **516**, 227–230 (2014).
- Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **397**, 601–604 (1999).
- de Gennes, P. G. Wetting: statics and dynamics. *Rev. Mod. Phys.* **57**, 827–863 (1985).
- Kwok, D. Y. & Neumann, A. W. Contact angle measurement and contact angle interpretation. *Adv. Colloid Interface Sci.* **81**, 167–249 (1999).
- Amirfazli, A., Hanig, S., Müller, A. & Neumann, A. W. Measurements of line tension for solid-liquid-vapor systems using drop size dependence of contact angles and its correlation with solid-liquid interfacial tension. *Langmuir* **16**, 2024–2031 (2000).
- Schneemilch, M., Welters, W. J. J., Hayes, R. A. & Ralston, J. Electrically induced changes in dynamic wettability. *Langmuir* **16**, 2924–2927 (2000).
- Manukyan, G., Oh, J. M., van den Ende, D., Lammertink, R. G. H. & Mugele, F. Electrical switching of wetting states on superhydrophobic surfaces: a route towards reversible Cassie-to-Wenzel transitions. *Phys. Rev. Lett.* **106**, 014501 (2011).
- Schrader, M. E. Young–Dupré revisited. *Langmuir* **11**, 3585–3589 (1995).
- Iannuzzi, M. *et al.* Site-selective adsorption of phthalocyanine on h-BN/Rh(111) nanomesh. *Phys. Chem. Chem. Phys.* **16**, 12374–12384 (2014).
- Ma, H. *et al.* Chiral distortion of confined ice oligomers ($n = 5, 6$). *Langmuir* **28**, 15246–15250 (2012).
- Ding, Y., Iannuzzi, M. & Hutter, J. Investigation of boron nitride nanomesh interacting with water. *J. Phys. Chem. C* **115**, 13685–13692 (2011).
- Holzschuh, E., Fritsch, M. & Kündig, W. Measurement of the electron neutrino mass from tritium β -decay. *Phys. Lett. B* **287**, 381–388 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge financial support by the Swiss National Science Foundation within the funding instrument 'Sinergia'. S.F.L.M. acknowledges the receipt of an FP7 Marie Curie European reintegration grant, ERC grant OxideSurfaces and, together with S.D.F., support by FWO-Vlaanderen. We thank M. Schmid for help with the depiction of the STM images, G. Schütz for discussions and A. P. Seitsonen for the artwork in Fig. 1g.

Author Contributions T.G. conceived the project together with S.F.L.M., who designed and performed the electrochemical experiments, *in situ* STM and *in situ* contact angle measurements. A.H. prepared the nanomesh samples, analysed contact angle data and performed the TDS measurements. S.M. built the contact angle apparatus. O.G. drew our attention to dynamic contact angle measurements. S.D.F. and J.O. managed the Sinergia project. All authors contributed to discussions. S.F.L.M., A.H. and T.G. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.F.L.M. (stmerten@gmail.com) or T.G. (greber@physik.uzh.ch).

METHODS

Experimental details. The experiments were performed in a clean room on single-layer h-BN grown on 150-nm-thick single-crystalline rhodium films on Si(111) wafers, which were protected with a 40-nm yttria-stabilized zirconia (YSZ) diffusion barrier³³. The wetting angles were determined by a laboratory-built CCD camera set up to image at 15 frames per second and with an automated angle evaluation scheme (see Methods section 'Error analysis'). Electrochemical potential control was achieved using a Metrohm–Autolab PGSTAT101 potentiostat, using a laboratory-built glass syringe carrying a Ag/AgCl/3 M NaCl reference and Pt wire counter electrode. All electrochemical potentials are reported versus the Ag/AgCl reference. All glassware, including the glass capillary for *in situ* contact angle measurements, was cleaned by boiling in 20% nitric acid and rinsing with ultrapure water (MilliQ, Millipore, 18.2 MΩ cm), to ensure perfect wetting. The electrolyte was prepared from reagent-grade 70% HClO₄ (Sigma–Aldrich) and ultrapure water. For the deuterium desorption experiments, light water was replaced with D₂O (Fluka, D > 99.8%). All operations were carried out in an Ar-filled glove box in a clean room³³. *In situ* STM was performed with an Agilent PicoLE system, using an electrochemically etched W tip that was coated with a thermoplastic polymer to minimize the faradaic current. The thermal desorption experiments were performed in an apparatus equipped with a calibrated quadrupole mass spectrometer (QMS 200 M2, Pfeiffer Vacuum), and pumped with a turbomolecular pump (TMU 071, Pfeiffer Vacuum)¹⁷.

Thermal and chemical stability of the h-BN/Rh nanomesh. Experience in our laboratories since the discovery of the boron nitride nanomesh³ indicates a shelf life of several years under ambient conditions, stability on exposure to, and ultrasonication in, common solvents (both polar and nonpolar), and after 15-min cycles in commercial UV–ozone cleaners. Furthermore, the nanomesh can be heated to 400 °C in air and close to 1,000 °C in vacuum⁴. By contrast, when exposed to hot, strongly alkaline solutions, the h-BN appears to be removed, in line with observations for bulk h-BN (ref. 34). Low-energy Ar⁺ bombardment followed by annealing has been shown to produce atomically precise defects³⁵. Hydrogen intercalation appears not to adversely affect the stability of the boron nitride layer¹⁷. As shown here, electrochemical cycling can be performed many times without signs of deterioration (Fig. 1b). Emersion of the intercalated sample and transfer to vacuum are possible and yield similar amounts of intercalated hydrogen as in an all-vacuum experiment (next section, Fig. 2 and ref. 17).

Adventitious adsorption from solution or the laboratory atmosphere was not observed, and attempts to adsorb organic molecules from solution were unsuccessful. These observations, together with the widely available cleaning methods listed above, make the nanomesh amenable to experiments under standard laboratory conditions.

Ambient-to-ultrahigh-vacuum transfer and TDS. A critical factor in our ambient-to-vacuum experiment is the successful emersion of the sample from the electrochemical environment, which means that potential control is required until the solid–bulk electrolyte interface ceases to exist. This way, the h-BN/Rh–electrolyte system is never allowed to assume its open circuit potential, where electrochemical deintercalation of the hydrogen may occur. This also explains why we can rinse the sample with ultrapure water to remove superficially adsorbed deuterium and traces of perchloric acid, before transferring the sample to vacuum. The successful extraction of the sample from the electrolyte is aided by its mild intrinsic hydrophobicity, akin to previous work³⁶ in which hydrophobic adlayers on noble metal electrodes after transfer to vacuum were studied.

The sample mounting and transfer into the vacuum chamber took about 10 min, during which the sample was exposed to water and ambient air. Because we expected that the remaining solvent molecules on the sample may still contribute to a deuterium signal during the desorption experiment, the sample was rinsed with ultrapure water after removal from the D₂O solution and before entry into the vacuum system. Artefacts from this procedure were excluded by comparing TDS spectra from two samples that were handled identically, but of which only one was held at the potential required for intercalation (see Fig. 2).

In the design of the sample holder, special precautions were taken to limit heating and desorption as much as possible to the sample itself, and to avoid a background covering the deuterium signal from the surface or risking the chamber pressure exceeding the safe operation limit of the mass spectrometer channel electron multiplier. To maximize the signal-to-noise ratio, heating should occur as locally and as fast as possible. Extended Data Fig. 1a shows the sample holders used, which allow the sample to be annealed via resistive heating solely through the 150-nm-thick Rh film³³. In this configuration, only the top tungsten clamps became as hot as the sample and temperatures of 580 °C could be reached with a maximum pressure load of about 1.2×10^{-6} mbar. The sample configuration requires the use of a pyrometer to assess the sample temperature, which was

calibrated against a thermocouple reading of a Rh sample as shown in Extended Data Fig. 1b. Because the lowest measureable temperature by the pyrometer is around 350 °C, additional extrapolation of the curve sections outside this region was required.

Extended Data Fig. 1b displays all relevant information of a desorption experiment, such as the applied power to the sample, pyrometer, mass spectrometer and integral pressure readings. To determine the sample temperature, the following differential equations were used, which consider the power applied to the sample and its heat loss through radiation and thermal conduction to the sample holder:

$$\frac{dT}{dt} = aP(t) + bP(t)T(t) \quad (1)$$

$$\frac{dT}{dt} = aP(t) + bT(t) \quad (2)$$

with applied power P , sample temperature T , time t and fitting parameters a and b . Best results were obtained using equation (1) for the ascending temperature region and equation (2) for the descending temperature region of the heating/cooling curve. The extrapolated temperature was finally a complete numerical simulation of the temperature evolution, with room temperature as the starting point and the applied power being the only input variables.

Rate of hydrogen (de)intercalation. The hydrogen intercalation we study is closely related to underpotential deposition of hydrogen—that is, electrosorption at potentials positive from the hydrogen evolution equilibrium potential—which is known to occur on the platinum group metals³⁷, and to involve comparable energies on Pt and Rh^{38,39}. As shown in Fig. 1a, hydrogen intercalation into h-BN/Rh occurs at a potential that is about 50 mV less negative than that for underpotential deposition of hydrogen on bare Rh(111), which may be due to the fact that intercalating hydrogen does not have to compete with ClO₄[−] for adsorption on Rh after passing the boron nitride overlayer¹⁹.

From the *in situ* STM image in Fig. 1e, which was recorded immediately after switching the substrate potential from a value at which intercalated hydrogen is stable to a value at which electrochemical deintercalation takes place, it is apparent that recovery of the corrugation requires about half of the time taken to record one image. If the local corrugation is taken as a measure of deintercalation, then the line speed (8.1 lines per second) and image resolution (512 × 512 pixels) indicate a characteristic time of approximately 30 s. The forward process (intercalation) reproducibly led to a flattening in EC-STM in less than 1 s (Extended Data Fig. 2), indicating that intercalation is considerably faster than deintercalation. Restoration of the corrugation must be considered as the last step in the electrochemical deintercalation; it is not obvious which is the rate-determining process, to which the pore and wire areas of the nanomesh may contribute differently. The observed difference in the rate of intercalation and deintercalation may originate from unilateral catalytic activation by the Rh substrate²¹—that is, the different proton affinity of the h-BN/Rh and the h-BN/electrolyte interface—and is also convoluted with the kinetics of underpotential deposition of hydrogen⁴⁰.

Error analysis. The contact angles from one image frame are determined on the basis of fits of 32 sigmoid functions to the change in optical contrast for the left and right periphery of the drop, which are used to determine the location of the edge of the liquid drop. These fits, which are based on 60 raw data pixels, are all located in an interval of 0.25 mm centred at the footprint of the drop, and yield a fitting error for the edge location of about 0.8 μm. This error is five times less than the optical pixel width of the image and can be explained by the very sharp optical contrast between the background and the liquid drop. The angle is then fitted with two separate linear functions, one for the real edge between the background and the drop and one for its mirror image. The crossing of these two lines reveals the location of the footprint; the contact angle α is half the angle between the two fitted lines:

$$\alpha = 0.5[\arctan(|b_1|) + \arctan(|b_2|)] \quad (3)$$

in which b_1 and b_2 are the slopes of the two linear fits (real angle and mirror angle). Equation (3) is valid for the left and right angles. Gaussian error propagation of equation (3) yields:

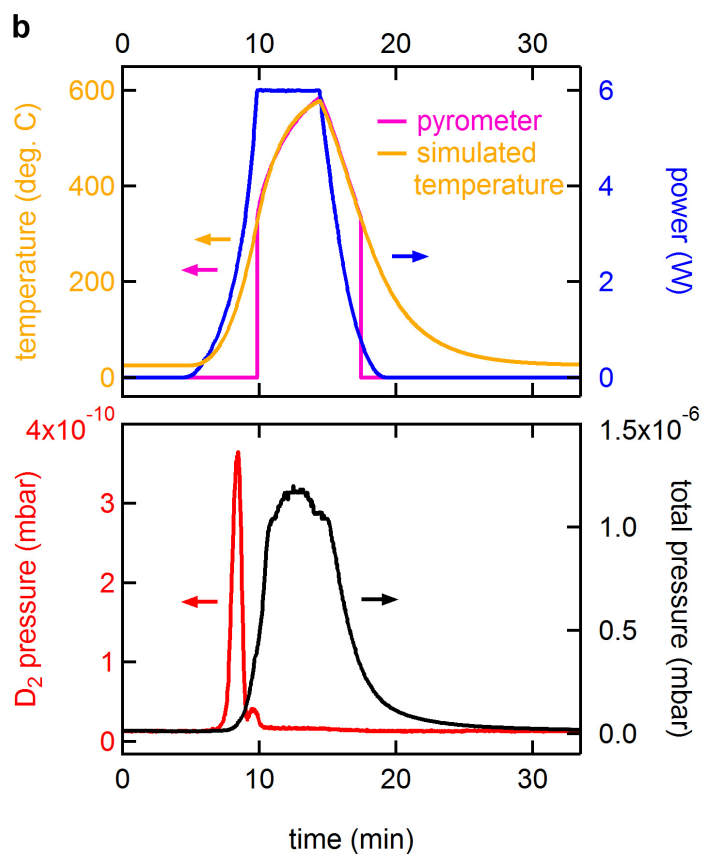
$$\Delta\alpha = \sqrt{\left(\frac{0.5\sigma b_1}{1+b_1^2}\right)^2 + \left(\frac{0.5\sigma b_2}{1+b_2^2}\right)^2}$$

in which σb_i is the error of the slope fits, and the accuracy of one measured contact angle is determined to be about 0.3°. For each extrapolation of the wetting angles

to $1/r=0$, we used about 400 angles, which were determined using the procedure outlined above. The four linear fits shown in Fig. 3c are based on 412, 410, 344 and 402 individual angle data points, respectively from top to bottom. Because of the high accuracy and the large number of data points, we may neglect the error of the individual angles. The influence of, for example, residual defects, which induce asymmetries between the left and right angles and angle spikes (Fig. 3a–c), is visibly larger. The error of the reduced contact angle is the fitting error of the θ -axis intersection; the largest error for the datasets in Fig. 3c is 0.8° , which is well below the size of the printed dot.

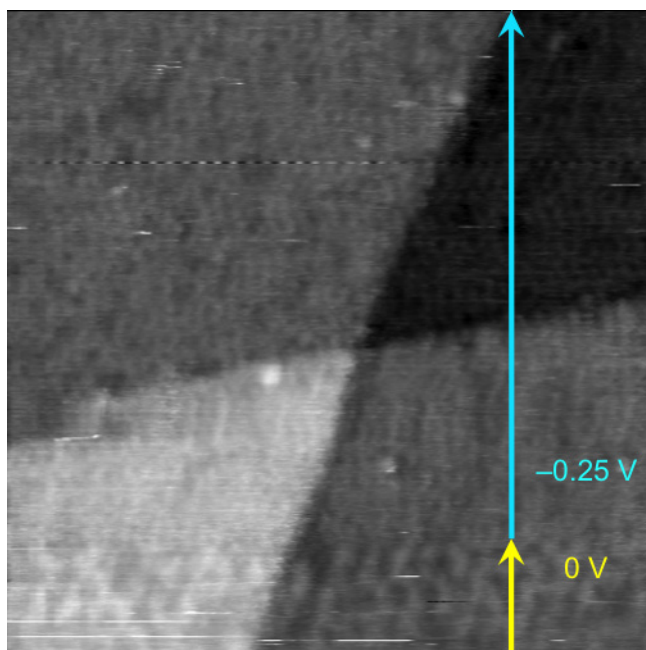
In Fig. 3d we show all experiments and the corresponding advancing and receding angles performed on three different samples. The largest errors are not from the angle determination procedure described above, but mainly from differences between the samples, sample handling and manual operation of the syringe. The uncertainties that we cite are derived from the standard deviations of the mean values of the advancing and receding angles in Fig. 3d, where advancing and receding angles are taken from the sigmoid fits at the same applied potential. Values of θ_a and θ_r of $76.0^\circ \pm 1.2^\circ$ and $24.5^\circ \pm 1.4^\circ$, respectively, for the corrugated and $84.3^\circ \pm 2.3^\circ$ and $40.1^\circ \pm 3.2^\circ$, respectively, for the flat state are found. The values and error bars for θ_0 , w_{LS} and γ_L are determined from the four angles and their standard deviations using Gaussian error propagation.

33. Hemmi, A. *et al.* High quality single atomic layer deposition of hexagonal boron nitride on single crystalline Rh(111) four-inch wafers. *Rev. Sci. Instrum.* **85**, 035101 (2014).
34. King, S. W., Nemanich, R. J. & Davis, R. F. Cleaning of pyrolytic hexagonal boron nitride surfaces. *Surf. Interface Anal.* **47**, 798–803 (2015).
35. Cun, H., Iannuzzi, M., Hemmi, A., Osterwalder, J. & Greber, T. Two-nanometer voids in single-layer hexagonal boron nitride: formation via the “can-opener” effect and annihilation by self-healing. *ACS Nano* **8**, 7423–7431 (2014).
36. Hansen, W. N. & Kolb, D. M. The work function of emersed electrodes. *J. Electroanal. Chem.* **100**, 493–500 (1979).
37. Koper, M. T. M. Blank voltammetry of hexagonal surfaces of Pt-group metal electrodes: comparison to density functional theory calculations and ultra-high vacuum experiments on water dissociation. *Electrochim. Acta* **56**, 10645–10651 (2011).
38. Jerkiewicz, G. & Zolfaghari, A. Determination of the energy of the metal–underpotential-deposited hydrogen bond for rhodium electrodes. *J. Phys. Chem.* **100**, 8454–8461 (1996).
39. Zolfaghari, A., Chayer, M. & Jerkiewicz, G. Energetics of the underpotential deposition of hydrogen on platinum electrodes. I. Absence of coadsorbed species. *J. Electrochem. Soc.* **144**, 3034–3041 (1997).
40. Łosiewicz, B., Jurczakowski, R. & Lasia, A. Kinetics of hydrogen underpotential deposition at polycrystalline rhodium in acidic solutions. *Electrochim. Acta* **56**, 5746–5753 (2011).



Extended Data Figure 1 | Thermal desorption spectroscopy. **a**, h-BN/Rh(111) thin-film sample and sample holder ready for the desorption experiment in the ultrahigh-vacuum chamber. **b**, Thermal desorption experiment with temperature and power data shown in the top panel and pressure data in the bottom panel. The purple data are pyrometer readings,

which were calibrated to a previous thermocouple measurement using a Rh crystal. The simulated temperature is a solution of equations (1) and (2) with the applied power and a starting temperature of 25 °C as input variables.



Extended Data Figure 2 | Electrochemical STM during hydrogen intercalation. The substrate potential during scanning (image scanned from bottom to top) was switched from $E_1 = 0\text{ V}$ (deintercalated, yellow arrow) to $E_2 = -0.25\text{ V}$ (intercalated, light blue arrow) at about one-fifth of the way from the lower edge of the STM image. On the timescale of imaging, intercalation-induced flattening of the surface is instantaneous, in sharp contrast to deintercalation (compare with Fig. 1e). Image size, $66\text{ nm} \times 66\text{ nm}$; tunnelling current, 0.1 nA ; tip potential fixed at -0.45 V .

Seasonality of temperate forest photosynthesis and daytime respiration

R. Wehr¹, J. W. Munger², J. B. McManus³, D. D. Nelson³, M. S. Zahniser³, E. A. Davidson⁴, S. C. Wofsy² & S. R. Saleska¹

Terrestrial ecosystems currently offset one-quarter of anthropogenic carbon dioxide (CO₂) emissions because of a slight imbalance between global terrestrial photosynthesis and respiration¹. Understanding what controls these two biological fluxes is therefore crucial to predicting climate change². Yet there is no way of directly measuring the photosynthesis or daytime respiration of a whole ecosystem of interacting organisms; instead, these fluxes are generally inferred from measurements of net ecosystem-atmosphere CO₂ exchange (NEE), in a way that is based on assumed ecosystem-scale responses to the environment. The consequent view of temperate deciduous forests (an important CO₂ sink) is that, first, ecosystem respiration is greater during the day than at night; and second, ecosystem photosynthetic light-use efficiency peaks after leaf expansion in spring and then declines³, presumably because of leaf ageing or water stress. This view has underlain the development of terrestrial biosphere models used in climate prediction^{4,5} and of remote sensing indices of global biosphere productivity^{5,6}. Here, we use new isotopic instrumentation⁷ to determine ecosystem photosynthesis and daytime respiration⁸ in a temperate deciduous forest over a three-year period. We find that ecosystem respiration is lower during the day than at night—the first robust evidence of the inhibition of leaf respiration by light^{9–11} at the ecosystem scale. Because they do not capture this effect, standard approaches^{12,13} overestimate ecosystem photosynthesis and daytime respiration in the first half of the growing season at our site, and inaccurately portray ecosystem photosynthetic light-use efficiency. These findings revise our understanding of forest-atmosphere carbon exchange, and provide a basis for investigating how leaf-level physiological dynamics manifest at the canopy scale in other ecosystems.

Much of what has been inferred about the behaviour of ecosystem photosynthesis, or 'gross ecosystem production' (GEP, defined as ecosystem-scale photosynthesis minus photorespiration), and 'daytime ecosystem respiration' (DER) in forests derives from eddy covariance measurements of their difference, denoted 'net ecosystem exchange' (NEE). NEE measurements have greatly advanced our understanding of carbon-cycle processes in terrestrial ecosystems, but the behaviours of GEP and DER have remained uncertain because eddy covariance does not distinguish one process from the other. In standard practice, a hypothesized response of GEP and/or DER to light, water, and/or temperature is used to make that distinction and thereby partition NEE into GEP and DER. For example, the oldest, simplest, and still most commonly adopted hypothesis is that DER follows the same function of air or soil temperature as does night-time ecosystem respiration^{12,13}, which is directly observable as night-time NEE. Another common partitioning hypothesis is that DER follows a function of air temperature of the same form found to apply to night-time NEE (but not necessarily with the same parameter values), while GEP follows a saturating function of

photosynthetically active radiation (PAR) of the same form found to apply to individual leaves¹⁴.

To date there has been no means of testing these partitioning hypotheses, but they are nevertheless used in hundreds of studies each year. The most popular partitioning algorithm¹³ alone has been cited more than 800 times since its debut in 2005, and similar methods have been explored since the onset of long-term eddy covariance measurements¹². The patterns and environmental responses of GEP and DER obtained from such methods have been used to design and evaluate terrestrial biosphere models for estimating large-scale biosphere-atmosphere interactions and predicting climate change^{4,5}. Partitioning has also been applied to a range of ecosystems to evaluate various remote sensing indices that may enable aircraft and satellite measurements of regional and global biosphere productivity but have contrasting seasonal patterns^{5,6,15,16}. And partitioning has been used to investigate ecosystem light-use efficiency and water-use efficiency, and to evaluate related production efficiency models intended to estimate regional and global biosphere productivity^{17,18}.

The general pattern observed via standard partitioning of NEE in temperate deciduous forests is that DER peaks or plateaus shortly after leaf expansion^{3,19}, one to two months ahead of the peak in belowground respiration¹⁹. Meanwhile, the ecosystem photosynthetic light-use efficiency peaks shortly after leaf expansion and then gradually declines through the growing season—as might occur because of leaf ageing or water stress—until autumnal senescence³.

To investigate the behaviour of GEP and DER, we partitioned three growing seasons of subhourly NEE measurements at the Harvard Forest into GEP and DER, on the basis of simultaneous eddy covariance measurements of the stable carbon isotopic composition (that is, ¹³C/¹²C) of NEE⁷. Our isotopic flux partitioning (IFP) algorithm (see Methods) exploits the fact that GEP and DER have ¹³C/¹²C signatures⁸ that are almost always distinguishable in subhourly NEE given the high precision (Extended Data Fig. 1) of our recently developed infrared laser spectrometer⁷ (see also Extended Data Fig. 2). We focus on comparisons with the most common standard partitioning algorithm¹³; comparisons with the partitioning method that incorporates a photosynthetic light-response function¹⁴ are broadly similar and can be found in the Methods.

Our analysis indicates that daytime ecosystem respiration differed fundamentally from standard predictions that were based on night-time NEE and temperature¹³ (Fig. 1): DER was only about half as large as night-time NEE in the first half of the growing season (June–July), but was roughly equal to night-time NEE in the second half (August–September). As belowground respiration typically varies by less than 10% between daytime and night-time at this site²⁰, the large discrepancy between daytime and night-time ecosystem respiration in the first half of the growing season suggests inhibition of leaf respiration by light, known as the Kok effect⁹. Such inhibition has been found to occur at the leaf level in many plant species¹⁰, including tree species¹¹

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. ²School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. ³Aerodyne Research Inc., Billerica, Massachusetts 01821, USA. ⁴Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, Maryland 21532, USA.

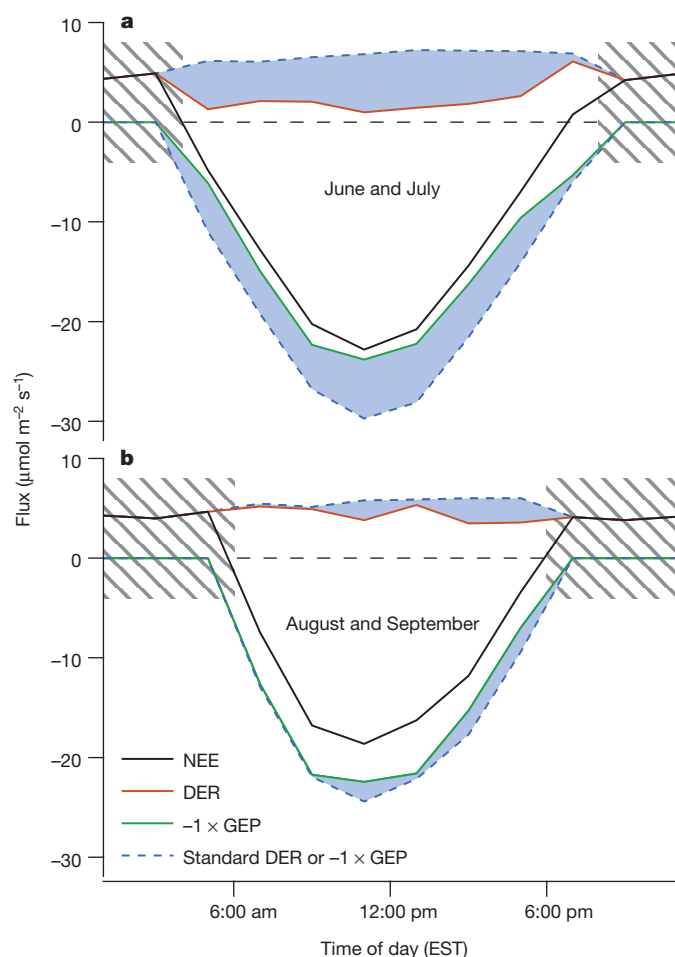


Figure 1 | Composite diel cycles show that photosynthesis and daytime respiration at the Harvard Forest are less than predicted in the first half of the growing season. **a, b,** Fluxes in the tower's relatively homogeneous and well sampled southwest quadrant, averaged across the three years, 2011–2013, for June to July (**a**) and August to September (**b**). Differences from the results of standard partitioning for the same data are shaded in blue. Lines connect means for each 2-hour bin ($20 \leq n \leq 94$). Partitioning is done for daylight periods only; GEP is set to zero in the dark (hatched areas).

(although we know of no published studies in red oak, which dominates our site). We therefore asked: what would cause inhibition of leaf respiration to occur during the first half of the growing season only, and is the magnitude of the discrepancy consistent with the overall respiration budget?

Previous work¹⁹ at this site on the seasonal patterns of aboveground and belowground respiration—a multiyear synthesis of more than 100,000 flux tower and soil chamber observations—suggests that in June, at night, aboveground respiration typically accounts for more than 50% of total ecosystem respiration, but that this proportion declines to about 10% in August (Fig. 2b, gold line). June–July is the period in which leaves are still thickening after expansion^{15,16} and are presumably continuing growth respiration associated with that thickening, which might explain the elevated night-time respiration of the canopy. (In contrast, night-time canopy respiration is roughly stable from June through to September at a nearby conifer-dominated site¹⁹.) We infer that aboveground respiration is strongly inhibited by light during the day at our site, causing DER to be about half as large as night-time ecosystem respiration in June–July but almost equal to night-time ecosystem respiration in August–September (Fig. 1). Near-complete inhibition of leaf respiration by light has been reported for individual leaves of some species¹⁰, and so both the seasonal pattern and the magnitude of the discrepancy between daytime and night-time

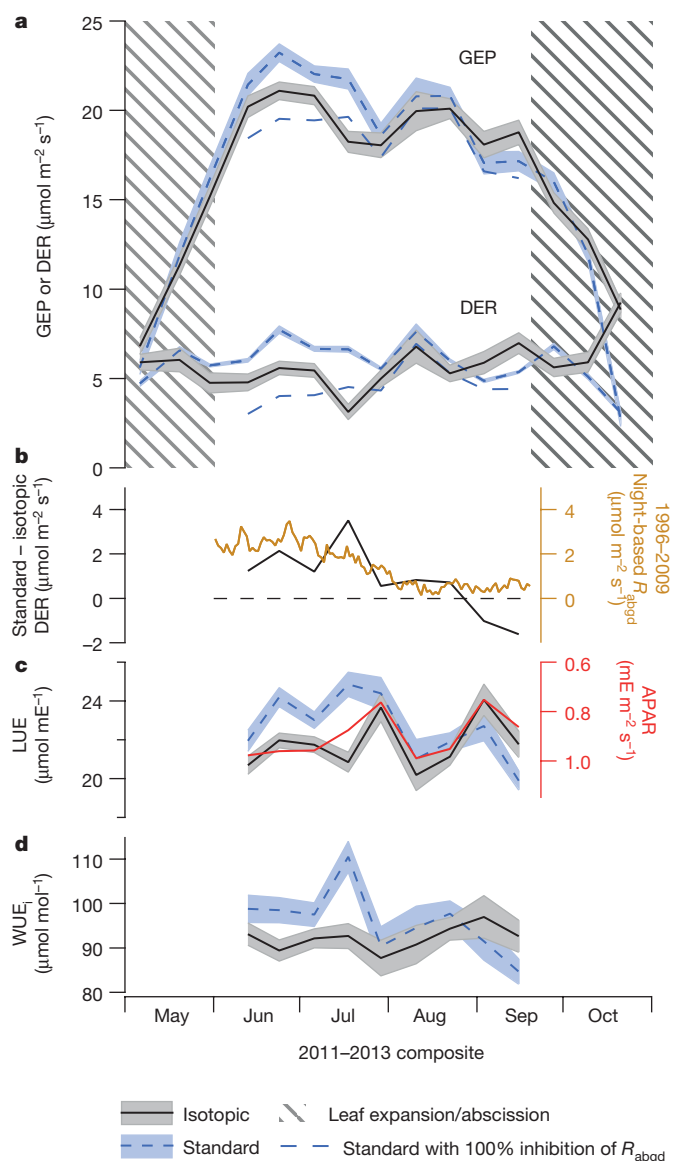


Figure 2 | Composite seasonal cycles of GEP and DER indicate strong inhibition of aboveground respiration by light and sustained photosynthetic efficiency. Results from isotopic partitioning, standard partitioning, and standard partitioning adjusted for 100% inhibition of aboveground respiration by light (see Methods), across all forest quadrants. **a,** GEP and DER. **b,** Discrepancy between standard and isotopic partitioning (black line), with the gold line showing the 1996–2009 mean seasonal pattern of aboveground respiration (R_{abgd}) estimated from soil chambers and night-time NEE¹⁹. **c,** Light-use efficiency (LUE; isotopic and standard partitioning), with absorbed photosynthetically active radiation (APAR) inverted in red. **d,** Intrinsic water-use efficiency (WUE_i). Lines connect means over all daylight hours for each 12-day bin; pale bands show standard errors of the means calculated from variability within each bin ($64 \leq n \leq 431$). Bands are omitted from the 100%-inhibition lines for clarity. Hatched areas indicate periods of leaf expansion and abscission.

ecosystem respiration are consistent with a plausible ecosystem-scale Kok effect.

Standard partitioning calculates DER from night-time NEE without accounting for any inhibition of respiration by light, and so the seasonally varying discrepancy between night-time and daytime ecosystem respiration (as determined by isotopic partitioning) corresponds to a similar, seasonally varying discrepancy between standard and isotopic estimates of DER (Fig. 2a, b). However, standard partitioning—and therefore the discrepancy between standard and isotopic partitioning—can be complicated by horizontal heterogeneity in ecosystem

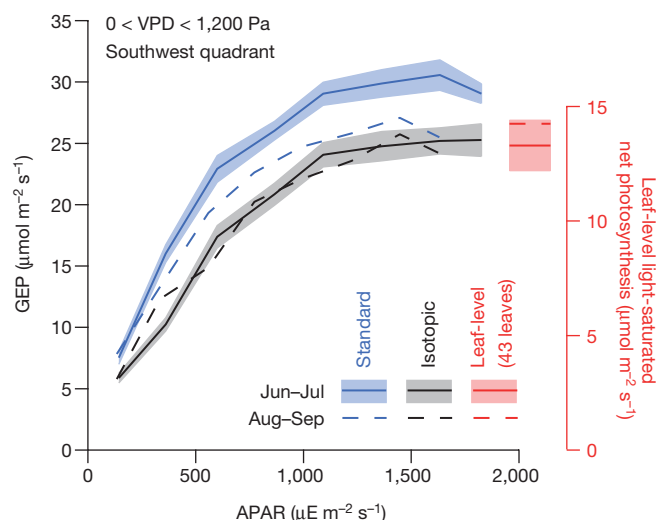


Figure 3 | The ecosystem-scale light-response curve is invariant over the season. Responses of GEP (from isotopic and standard partitioning) to APAR, in June–July and August–September (before leaf abscission), averaged across all three years, for vapour pressure deficits (VPDs) of between 0 Pa and 1,200 Pa, in the flux tower's southwest quadrant. Lines connect means for each APAR bin, and pale bands show standard errors of the means calculated from variability within each bin ($14 \leq n \leq 63$). Also shown (right-hand axis) are the leaf-level light-saturated net photosynthesis rates (that is, net leaf CO_2 -uptake rates) reported for the years 1991 to 1992 in ref. 26, based on 43 red oak leaves in mid July and 44 red oak leaves in late August; the small increase from mid July to late August was also reported for red maple leaves in ref. 26. Bands are omitted from the August–September curves for clarity.

respiration surrounding the flux measurement tower. Standard partitioning uses all night-time fluxes in a 4- or 6-day time window to predict DER for each individual daytime flux measurement. The night-time fluxes typically correspond to many different sectors of the forest, according to the winds; but the individual daytime flux being partitioned corresponds to one particular sector. Therefore, standard partitioning effectively averages across sectors, underestimating DER for high-respiration sectors and overestimating it for low-respiration sectors⁸. Moreover, if the winds differ systematically between day and night, standard DER predictions can be further biased. At our site, ecosystem respiration was more than two times larger to the north-west of the flux measurement tower than it was to the southwest⁸, and the day-versus-night wind biases in our data set were such that the June–July discrepancy between standard and isotopic estimates of DER is understated when averaging across all sectors of the forest (as in Fig. 2), but exaggerated when averaging only within the southwest quadrant (as in Fig. 1). Without these sampling biases, the standard method would have overestimated DER by about 100% in June–July (see Methods).

We have thus far described only the period between the completion of leaf expansion and the onset of autumnal senescence (June–September). Prior to leaf expansion, aboveground respiration already constitutes more than half of night-time ecosystem respiration, probably because of respiration associated with bud and leaf development, branch elongation, and wood production¹⁹. In this period, we found that daytime and night-time ecosystem respiration were roughly equal, so that the standard and isotopic partitioning methods agreed (Fig. 2a). This result is expected, as respiration should be inhibited by light only in photosynthesizing tissues.

The revised seasonal pattern in DER corresponds to a new seasonal pattern in GEP, because DER and GEP are constrained to sum to the measured NEE in both partitioning methods. Whereas standard partitioning suggested a gradual decline in the response of GEP to absorbed photosynthetically active radiation (APAR) through the

growing season, isotopic partitioning showed that the response of GEP to APAR was stable between the completion of leaf expansion and the onset of autumnal senescence (Fig. 3). Neither GEP (Fig. 2a) nor the canopy light-use efficiency (LUE, being GEP/APAR) (Fig. 2c) determined by isotopic partitioning exhibited the pronounced early-season peak that typifies standard estimates at this site and in temperate forests more generally³. Compared with isotopic partitioning, and aside from the directional sampling biases mentioned above, standard partitioning overestimated GEP by about 25% in June–July.

Qualitatively, the observed saturating response of GEP to APAR (Fig. 3) and the resulting negative correlation between LUE and APAR (Fig. 2c) make physiological sense for three reasons. First, as the amount of light increases, the photosynthetic apparatus of a leaf becomes increasingly limited by other factors, such as CO_2 supply²¹. Second, cloudy or hazy conditions cause the amount of light to decline but the remaining light to be more diffuse and therefore spread more evenly among the canopy leaves, strongly increasing assimilation by otherwise shaded leaves but only mildly reducing assimilation by otherwise sunlit, light-saturated leaves²². Third, high-light conditions tend to be associated with drier air, which can lower LUE by inducing stomatal restriction of the CO_2 supply. Quantitatively, the observed seasonal-scale relationship between isotopically determined LUE and APAR ($r^2 = 0.75$, $P = 0.002$ between the completion of leaf expansion and the onset of abscission; Fig. 2c and Extended Data Fig. 3) is expected, in the sense that it is consistent with relationships obtained from both standard and isotopic partitioning at shorter timescales (Extended Data Fig. 4). In contrast, LUE predicted by standard partitioning did not yield any significant seasonal-scale correlation with APAR ($r^2 = 0.07$, $P = 0.39$) (Fig. 2c and Extended Data Fig. 3). There would indeed be such a correlation were it not for the June–July LUE peak, which we argue is an artefact arising because standard partitioning does not capture the inhibition of leaf respiration by light.

Another common metric of photosynthetic efficiency is canopy-intrinsic water-use efficiency (WUE_i)²³, which is the ratio of carbon gained by photosynthesis (GEP) to water lost by transpiration (E_T), controlling for the water-vapour mole fraction gradient between the inside of the leaves and the air (Δw). Thus, $\text{WUE}_i = 1.6 \times \Delta w \times \text{GEP}/E_T$ (the factor of 1.6 in this definition accounts for the fact that water vapour diffuses 1.6 times more quickly than CO_2). We found that WUE_i was more stable throughout the growing season than is predicted by standard partitioning (Fig. 2d), implying a stable relationship between canopy stomatal conductance and GEP.

In some forests, LUE is believed to decline before the onset of senescence owing to water stress²⁴. Some plants exhibit a decline in LUE as leaves age²⁵. In our forest, however, leaf-level gas-exchange measurements indicate little or no decline in photosynthetic capacity for our site-dominant tree species before senescence^{15,26} (Fig. 3, red lines). Moreover, our observations show that the soil water content increased, and the atmospheric vapour pressure deficit decreased, from July to September (Extended Data Fig. 5): that is, water stress was decreasing throughout that period. Thus, LUE should not be expected to decline before senescence at our site, in accordance with our partitioning results. Indeed, given existing literature on the leaf-level Kok effect^{9–11}, and on the seasonal patterns of aboveground respiration¹⁹ and leaf-level photosynthetic capacity at our site^{15,26}, the seasonal patterns of DER and GEP derived from isotopic partitioning appear more consistent with present knowledge than do those obtained by standard partitioning.

Our findings suggest a need to reappraise standard partitioning methods at other flux tower sites, in other ecosystems, to determine whether the patterns we have observed at the Harvard Forest are common elsewhere. For example, in equatorial Amazonian forests, standard eddy-covariance-based estimates of GEP and LUE increase throughout the dry season. That increase contradicts some model predictions, but is supported by observations of increased leaf flushing at the start of the dry season^{27,28}. Our findings, however, point to the possibility that

some of the apparent seasonality in tropical GEP might actually be a partitioning artefact introduced by not accounting for seasonality in (light-inhibited) leaf respiration, with young leaves respiring more (at night), as here.

The empirical understanding of carbon exchange between temperate forests and the atmosphere that underpins efforts to estimate global biosphere productivity^{5,6} and to predict climate change^{4,5} holds that ecosystem respiration is greater during the day than at night, and that the canopy photosynthetic light-use efficiency gradually declines throughout the growing season. But the strong apparent light-mediated inhibition of canopy respiration and the invariance of the canopy photosynthetic light response found here challenge that understanding. These phenomena also highlight the central role of leaf-level physiological dynamics in ecosystem-scale responses to the environment, and the potential benefit of incorporating more accurate representation of such dynamics into Earth-system models. Our study suggests that ecosystem-scale isotopic flux measurements could provide a general basis for exploring how leaf-level dynamics play out in ecosystems of varying composition and phenology around the world.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 September 2015; accepted 23 March 2016.

- Intergovernmental Panel on Climate Change. *Climate Change 2013: The Physical Science Basis* (Cambridge Univ. Press, 2013).
- Schimel, D. *et al.* Observing terrestrial ecosystems and the carbon cycle from space. *Glob. Change Biol.* **21**, 1762–1776 (2015).
- Falge, E. *et al.* Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agric. For. Meteorol.* **113**, 53–74 (2002).
- Stöckli, R. *et al.* Use of FLUXNET in the community land model development. *J. Geophys. Res.* **113**, 1–19 (2008).
- Parazoo, N. C. *et al.* Terrestrial gross primary production inferred from satellite fluorescence and vegetation models. *Glob. Change Biol.* **20**, 3103–3121 (2014).
- Turner, D. P. *et al.* Scaling gross primary production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sens. Environ.* **88**, 256–270 (2003).
- Wehr, R. *et al.* Long-term eddy covariance measurements of the isotopic composition of the ecosystem–atmosphere exchange of CO₂ in a temperate forest. *Agric. For. Meteorol.* **181**, 69–84 (2013).
- Wehr, R. & Saleska, S. R. An improved isotopic method for partitioning net ecosystem–atmosphere CO₂ exchange. *Agric. For. Meteorol.* **214–215**, 515–531 (2015).
- Kok, B. On the interrelation of respiration and photosynthesis in green plants. *Biochim. Biophys. Acta* **3**, 625–631 (1949).
- Heskel, M. A., Atkin, O. K., Turnbull, M. H. & Griffin, K. L. Bringing the Kok effect to light: a review on the integration of daytime respiration and net ecosystem exchange. *Ecosphere* **4**, 98 (2013).
- Sun, J. *et al.* Estimating daytime ecosystem respiration to improve estimates of gross primary production of a temperate forest. *PLoS One* **9**, <http://dx.doi.org/10.1371/journal.pone.0113512> (2014).
- Wofsy, S. C. *et al.* Net exchange of CO₂ in a mid-latitude forest. *Science* **260**, 1314–1317 (1993).
- Reichstein, M. *et al.* On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob. Change Biol.* **11**, 1424–1439 (2005).
- Lasslop, G. *et al.* Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Glob. Change Biol.* **16**, 187–208 (2010).
- Dillen, S. Y., Phillips, N., de Beeck, M. O., Hufkens, K. & Buonanduci, M. Agricultural and forest meteorology. *Agric. For. Meteorol.* **160**, 60–68 (2012).
- Keenan, T. F. *et al.* Tracking forest phenology and seasonal physiology using digital repeat photography: a critical assessment. *Ecol. Appl.* **24**, 1478–1489 (2014).
- Tang, X. *et al.* How is water-use efficiency of terrestrial ecosystems distributed and changing on Earth? *Sci. Rep.* **4**, 7483 (2014).
- Yuan, W. *et al.* Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuille database. *Agric. For. Meteorol.* **192–193**, 108–120 (2014).
- Giasson, M. A. *et al.* Soil respiration in a northeastern US temperate forest: a 22-year synthesis. *Ecosphere* **4**, 140 (2013).
- Savage, K., Davidson, E. A. & Tang, J. Diel patterns of autotrophic and heterotrophic respiration among phenological stages. *Glob. Change Biol.* **19**, 1151–1159 (2013).
- De Pury, D. G. G. & Farquhar, G. D. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant Cell Environ.* **20**, 537–557 (1997).
- Knohl, A. & Baldocchi, D. D. Effects of diffuse radiation on canopy gas exchange processes in a forest ecosystem. *J. Geophys. Res.* **113**, G02023 (2008).
- Beer, C. *et al.* Temporal and among-site variability of inherent water use efficiency at the ecosystem level. *Global Biogeochem. Cycles* **23**, GB2018 (2009).
- Grassi, G., Vicinelli, E., Ponti, F., Cantoni, L. & Magnani, F. Seasonal and interannual variability of photosynthetic capacity in relation to leaf nitrogen in a deciduous forest plantation in northern Italy. *Tree Physiol.* **25**, 349–360 (2005).
- Constable, G. A. & Rawson, H. M. Effect of leaf position, expansion and age on photosynthesis, transpiration and water use efficiency of cotton. *Funct. Plant Biol.* **7**, 89–100 (1980).
- Bassow, S. L. & Bazzaz, F. A. How environmental conditions affect canopy leaf-level photosynthesis in four deciduous tree species. *Ecology* **79**, 2660–2675 (1998).
- Restrepo-Coupe, N. *et al.* What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network. *Agric. For. Meteorol.* **182–183**, 128–144 (2013).
- Wu, J. *et al.* Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science* **351**, 972–976 (2016).

Acknowledgements This research was supported by the US Department Of Energy (DOE), Office of Science, Terrestrial Ecosystem Science (TES) program (award DE-SC0006741), and the Agnese Nelms Haury Program in Environment and Social Justice at the University of Arizona. Soil chamber data were acquired with the help of K. Savage at the Woods Hole Research Center, under the DOE award. The Harvard Forest Environmental Measurements Site infrastructure is a component of the Harvard Forest Long-Term Ecological Research (LTER) site, supported by the National Science Foundation (NSF), and is additionally supported by the DOE TES program. Below-canopy PAR, soil temperature and soil moisture data were provided by E. Nicoll at the Harvard Forest, supported by the NSF LTER program. OCS data and SIF data were provided by the authors of refs 40 and 41, respectively.

Author Contributions S.R.S. conceived the study. S.R.S. and R.W. designed the study. R.W., J.B.M., D.D.N. and M.S.Z. developed and maintained the spectrometer. R.W. and J.W.M. set up and maintained the instrumentation and conducted the measurements. R.W. analysed the data and wrote the manuscript. R.W. and S.R.S. led the interpretation of the results with input from J.W.M., S.C.W. and E.A.D. All authors contributed to editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.W. (rawehr@email.arizona.edu) or S.R.S. (saleska@email.arizona.edu).

METHODS

Data availability. The measurements used, as well as ancillary site data, are publicly available from the Harvard Forest Data Archive (<http://harvardforest.fas.harvard.edu/harvard-forest-data-archive>) under records HF209 and HF004. The measurements are also publicly available at <ftp://saleskalab.eebweb.arizona.edu/>, along with the analysis algorithms used.

Site description. The measurements used here were acquired between May 2011 and October 2013 at the Harvard Forest Environmental Measurements Site^{13,29}, which is situated in a mostly deciduous temperate forest dominated by red oak and red maple (with some hemlock and red pine) in Massachusetts, USA. Average seasonal cycles of key environmental and forest variables are shown in Extended Data Fig. 5.

Measurements. Acquisition of isotopic fluxes (described previously⁷) was via a recently developed quantum cascade laser spectrometer that measures the isotopic composition of atmospheric CO₂ with unprecedented accuracy and precision. The long-term (>3-year) reproducibility in 100-s measurements by this spectrometer is $\pm 0.1\text{‰}$ for $\delta^{13}\text{C}$ and $\pm 0.12\text{‰}$ for $\delta^{18}\text{O}$ (95% confidence interval) (Extended Data Fig. 1) The short term (<3-hour) precision in 100-s measurements—which is the relevant metric for distinguishing isotopic signatures in updrafts and down-drafts and, hence, for using the eddy covariance method—is $\pm 0.04\text{‰}$ for $\delta^{13}\text{C}$ and $\pm 0.06\text{‰}$ for $\delta^{18}\text{O}$ (95% confidence interval).

Environmental and forest variables used in this study (Extended Data Fig. 5) include:

Plant area index. This was measured optically across a plot array in the flux tower footprint throughout the growing season (as part of routine site operations), interpolated to our 40-minute time grid, and converted to leaf area index (LAI) as described elsewhere⁸. The period between the completion of leaf expansion and the onset of leaf abscission is defined as the time during which the LAI was greater than 95% of its stable summertime maximum.

Absorbed photosynthetically active radiation (APAR). This is calculated as the difference between above-canopy PAR and below-canopy PAR (that is, neglecting reflection), the latter of which was calculated as the mean of six measurements (using Li-Cor LI-190S quantum sensors mounted 1 metre above the ground) distributed within the flux tower footprint.

Soil temperature. This was calculated as the mean of measurements from eight copper-constantan thermocouples buried 10 cm below the ground surface, distributed within the typical flux tower footprint.

Volumetric soil moisture. This was measured for the depth range 0–30 cm at four locations in the typical flux tower footprint.

Isotopic flux partitioning. Since its original exposition³⁰, the isotopic partitioning of NEE has been tested in several ecosystems^{31–38} but has not been generally adopted owing to instrumental limitations. Motivated by the improvements in field instrumentation described above, we recently extended the theory of isotopic flux partitioning (IFP), and successfully demonstrated its application at the Harvard Forest for the first year of our measurements, quantifying its uncertainties and potential biases⁸. Here we briefly recapitulate our method.

The basic idea of IFP is to determine GEP and DER using their isotopic signatures and the magnitude and isotopic composition of their residual, NEE; in other words, to solve the following set of two mass balance equations for the two unknowns, DER and GEP, where δ_{NEE}^{13} , δ_{DER}^{13} , and δ_{GEP}^{13} describe the ratios (expressed relative to a standard material) of ^{13}C to ^{12}C in NEE, DER, and GEP, respectively:

$$\text{NEE} = \text{DER} - \text{GEP}$$

$$\delta_{\text{NEE}}^{13}\text{NEE} = \delta_{\text{DER}}^{13}\text{DER} - \delta_{\text{GEP}}^{13}\text{GEP}$$

Because the isotopic signature of GEP (that is, δ_{GEP}^{13}) varies substantially during the day, additional equations based on leaf anatomy and biochemistry are used to express the isotopic signature of GEP in terms of GEP itself, chiefly via the canopy-integrated stomatal conductance, determined from heat and water fluxes measured by eddy covariance⁸. In our recently extended theory of isotopic partitioning⁸, the two equations above are replaced by a system of six equations in which DER is broken down into foliar and non-foliar respiration, and GEP is broken down into gross photosynthesis and photorespiration. The breakdown is necessary because each of these four processes has its own isotopic signature.

The isotopic signature of non-foliar respiration is the only signature that is measured, by a combination of soil chamber and night-time Keeling plot measurements⁸. This method is weighted strongly to belowground respiration⁸. To the degree that the isotopic signature of non-foliar aboveground respiration differed from that of belowground respiration, our isotopic signature of overall non-foliar respiration would therefore have been in error. However, we showed previously that the partitioning is not very sensitive to that signature: a relatively

large 1‰ error in the signature of overall non-foliar respiration leads to an error of just 3% in GEP⁸.

Although our IFP algorithm⁸ treats foliar and non-foliar respiration separately because of their independent isotopic signatures, the rates of foliar and non-foliar respiration cannot be determined separately from one another with confidence by IFP (at least at present) because neither is directly measurable or predictably related to other variables—as, for example, photorespiration is related to photosynthesis by the photocompensation point (which is measurable) and DER is related to photosynthesis and photorespiration by NEE (which is measurable). In other words, we have enough information to solve the equations for DER but not to apportion DER to foliar and non-foliar sources. The attribution of patterns in DER to patterns in foliar respiration in the main text is therefore based instead on comparison to belowground respiration measured by soil chambers.

For the purposes of the calculation, we use a plausible *a priori* value for the rate of foliar respiration, and test how sensitive the GEP-versus-DER partitioning results are to this value. We find that the choice of *a priori* value, within plausible bounds, is inconsequential for the values of GEP and DER output by IFP (Extended Data Fig. 6), as was also found previously⁸. The partitioning of NEE into GEP and DER is thus robust even if IFP's internal foliar/non-foliar apportioning is not. The plausible bounds for daytime foliar respiration are approximately: (1) zero, and (2) the total night-time aboveground respiration, which was determined by a multiyear synthesis of flux tower and soil chamber data in the typical flux tower sampling footprint at this site¹⁹. For the partitioning in the main text, we set the *a priori* value in the middle of the range; that is, for each daytime flux measurement, we calculated leaf respiration as $0.5 \times p \times \text{NER}$, where NER is night-time ecosystem respiration (that is, night-time NEE) for the same wind direction, and *p* is the (seasonally varying) proportion of NER that is aboveground respiration according to the multiyear synthesis mentioned above. In this case, Extended Data Fig. 6 shows that the maximum possible associated error in GEP at any time of the growing season is 3.5%. Given the insensitivity of GEP and DER to the *a priori* rate of foliar respiration, and given that the consideration of foliar respiration in the main text is based on comparison of DER to other kinds of measurement rather than on IFP's internal apportioning, the *a priori* rate of foliar respiration does not affect our findings concerning the inhibition of foliar respiration (or any other findings).

With a measured signature of non-foliar respiration and an *a priori* rate of foliar respiration, the other six variables can be determined simultaneously by solution of the six equations. The equations neglect variation among the leaves within the canopy with regards to physiology and microenvironment, a source of uncertainty in the method that—as we argue in the section ‘Uncertainty in the seasonal patterns’ (below) and in Extended Data Fig. 7—is unlikely to significantly affect our findings.

Successful application of isotopic partitioning requires that the isotopic signatures of GEP and DER be distinct (otherwise the partitioning equations above are not independent). That is the case predominantly because of the large (4‰) diurnal variation in the signature of GEP and the small (<0.25‰) diurnal variation in the signature of DER. But we also see consistently distinct signatures even at seasonal timescales averaged over three years (Extended Data Fig. 2).

A minor alteration to the previous method⁸, enabled by our larger data set here (three growing seasons instead of one), concerns the empirical function used to interpolate eddy-covariance-based estimates of stomatal conductance into periods with significant evaporation (for example, from the soil surface), which would otherwise contaminate the stomatal conductance estimation because it is a water flux that does not pass through the stomata. Previously⁸, only one year of data was used. Here the function was fit using all three years of available data, and the resulting set of fitted parameters was still able to closely reproduce the measured half-hourly water flux ($r^2 = 0.92$) in selected evaporation-free periods⁸ that we used for testing the method. The strong ability to reproduce the water flux gives us high confidence that stomatal conductance values are well estimated during evaporative periods, and that IFP can therefore be extended to these times.

Note that some parameters used for isotopic partitioning (for example, mesophyll conductance) might be inaccurate during periods of leaf expansion (that is, increase in leaf area following budburst) or autumnal senescence (that is, coloration and abscission), shown as hatched areas in Fig. 2 and Extended Data Figs. 2 and 5–10.

Standard partitioning. For comparisons, we followed the standard partitioning algorithms exactly as described^{13,14}, except that we used slightly longer time windows for building regressions in order to substantially reduce the number of regressions that were based on only a few data points (we used 6-day windows where 4-day windows were prescribed, and 15-day windows where either 12- or 15-day windows were prescribed). All partitioning methods partitioned the same set of NEE measurements.

Sampling biases in standard partitioning. There are two competing sampling biases in the standard partitioning method used in the main text¹³, both of which are related to the fact that both night-time and daytime ecosystem respiration

(the latter as determined by IFP) were more than two times higher in areas to the northwest of the flux measurement tower than they were to the southwest⁸ (GEP from IFP was also higher to the northwest, by 28%). The Harvard Forest is not odd in this regard; directional variation in the flux measured by eddy covariance can result from true ecosystem heterogeneity or from correlation of environmental conditions with wind direction, caused by synoptic weather patterns.

The first and larger bias arises because the standard method calculates DER from all night-time flux measurements in a 4- or 6-day time window; thus standard partitioning of a given flux measurement, corresponding to a given wind direction, is influenced by fluxes from many wind directions. This bias leads the standard method to overestimate DER in the southwest quadrant and underestimate it in the northwest quadrant; however, the bias is negligible if estimates of DER from all quadrants are averaged (as in Fig. 2). Restricting the standard partitioning method to use night-time fluxes from only one quadrant in its regressions would require expansion of the time window to almost a month in order to maintain sufficient points for each regression, which would raise concerns about seasonal variation biasing the method¹³. The isotopic method, on the other hand, partitions each flux measurement independently and thereby captures the variation of DER and GEP between different areas of the forest⁸.

The second bias arises because natural wind patterns caused the night-time and daytime fluxes to be associated with different areas of the forest. In particular, in June and July (but not in August or September), the high-respiration areas to the northwest were sampled commonly during the daytime but rarely at night. This bias leads the standard method to underestimate DER in June and July but not in August or September. Following the discrepancy between night-time NEE and DER from IFP, the standard partitioning method should overestimate DER from IFP by about 100% in June–July. Instead, when estimates of DER from all quadrants are averaged to remove the first and larger bias (discussed above), this second bias reduces that overestimation to about 35%. If our forest were more homogeneous, the standard method would have overestimated DER by about 100%.

We include all quadrants in the averages in Fig. 2 because it is focused on comparing the two partitioning methods, and the second bias is smaller than the first. On the other hand, we restrict Figs. 1 and 3 to the relatively homogeneous southwest quadrant because they are focused on the day–night difference in respiration and on the light-response of GEP, which are more accurately portrayed without the variability associated with sampling different quadrants.

Standard partitioning adjusted for 100% inhibition of aboveground respiration by light. The dashed blue lines in Fig. 2 were obtained by scaling DER from standard partitioning by the seasonally varying ratio of belowground respiration to total ecosystem respiration in ref. 19, which was estimated from soil chambers and night-time NEE.

Comparison with partitioning based on light-response curves. Regression-based partitioning incorporating a photosynthetic light-response function in addition to a respiratory temperature-response function¹⁴, like the standard regression-based method discussed in the text¹³, overestimated DER from IFP in July; however, it diverged from IFP and the more standard method at other times of the year (Extended Data Fig. 8). At this site, there is large horizontal variation in ecosystem respiration, so that changes in the flux tower sampling footprint (that is, in wind speed or direction) can cause the measured respiration to vary by a factor of 3 or more⁸. The dual-function method can only interpret such variation as being driven by PAR, sometimes leading to unrealistic light-response curves and erratic results⁸; examples are evident in August and October in Extended Data Fig. 8. Also like the standard method discussed in the text, dual-function partitioning yielded a light-use efficiency that showed no significant correlation with APAR on the seasonal scale ($r^2 = 0.01$, $P = 0.85$) (Extended Data Figs. 3 and 8), although it did correlate with APAR on shorter timescales (Extended Data Fig. 9).

Uncertainty in the seasonal patterns. Because the estimated systematic uncertainty in IFP at this site exceeds 17% of GEP⁸ (1 standard error), we examined the potential for errors in IFP to produce the observed seasonal pattern of disagreement with standard partitioning. The two largest sources of uncertainty in IFP—the only two that could account for a 10% error in GEP⁸ in the period after leaf expansion and before senescence—are: (1) neglect of heterogeneity among the canopy leaves and their microenvironments, that is, the ‘big-leaf’ assumption, and (2) the prescribed isotopic fractionation by enzyme-catalysed fixation of CO₂, which is based on measurements in only a few non-forest species.

Error owing to the big-leaf assumption would depend on the physical and physiological structure of the canopy and on the distribution of the light source (that is, the angle of the sun and the diffuse light fraction). Because the canopy structure does not vary appreciably between full leaf expansion and the onset of senescence, we can minimize any error owing to the big-leaf assumption during that period by restricting our analysis to data acquired under fully diffuse

light—that is, when the light source was distributed uniformly over the sky. We found that the seasonal pattern of the disagreement between IFP and standard partitioning is not qualitatively changed by restricting the analysis to fully diffuse light conditions (diffuse light fraction > 0.9), or by systematic error in the isotopic fractionation by enzyme-catalysed fixation of CO₂ (aside from a shift in the mean disagreement; Extended Data Fig. 7). Thus the seasonal pattern of disagreement between IFP and standard partitioning is unlikely to be an artefact of uncertainty in IFP.

Random errors. The standard error bands in Figs. 2 and 3 are based on total variability within each averaging bin and therefore represent an upper bound on the effect of random measurement error. The random error in the isotopic partitioning of individual NEE measurements has previously been shown⁸ to be negligible compared with the random error in the eddy covariance measurement of NEE itself.

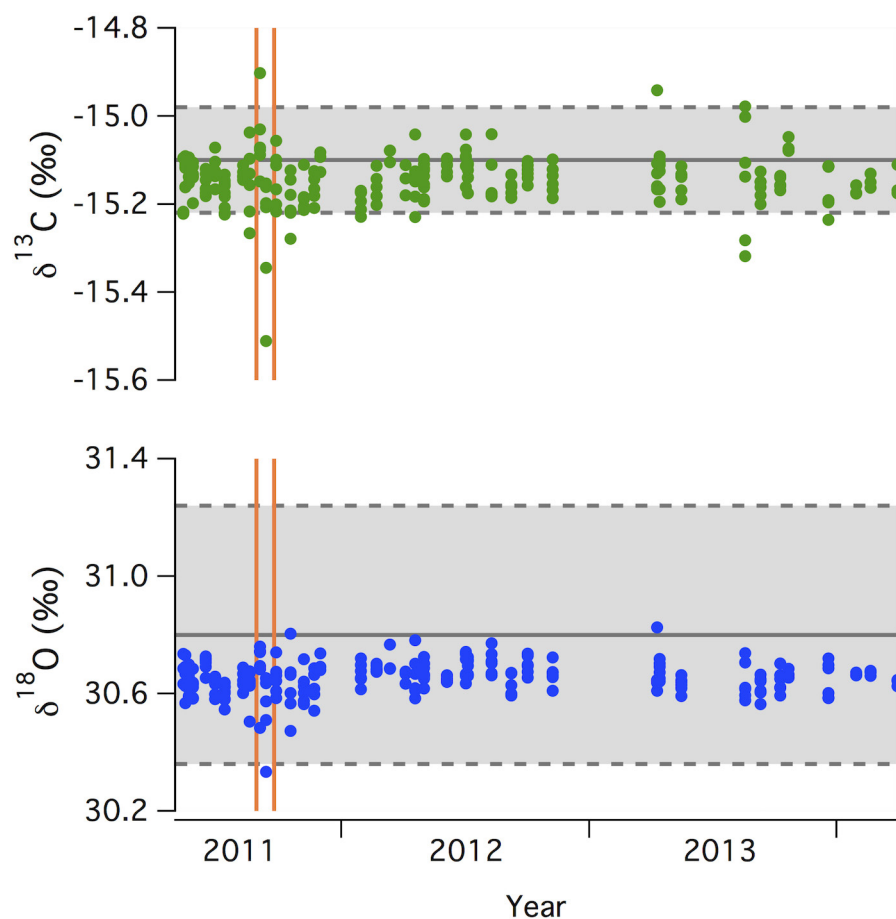
Ogée *et al.*³⁹ once argued, using Bayesian methods, that even with improved instrumentation, the random error in isotopic partitioning using ¹³C would remain so large that the partitioning would be useful only when averaging over many individual flux measurements. The rationale for the large error estimate was that the typical isotopic disequilibrium (~2‰) is not much larger than the random measurement error in the isotopic signatures of photosynthesis and respiration, and in the isotopic composition of NEE. However, Ogée *et al.*'s analysis did not account for correlations between the errors in these various isotopic quantities. Because the photosynthetic signature is determined by the partitioning equations (as opposed to being measured), random error in the photosynthetic signature derives entirely from, and is therefore correlated with, random error in the measured variables (especially in the isotopic composition of NEE). That correlation reduces the measurement-derived random error in DER and GEP relative to Ogée *et al.*'s estimates³⁹, as confirmed by the actual variability in the retrieved gross fluxes not only here but also in all previous IFP studies^{31–39}. That variability demonstrates that ¹³C measurements are sufficiently precise to partition NEE measurements robustly on an hourly basis.

The effect of error in the isotopic composition of NEE on error in the photosynthetic signature using our equations was determined by sensitivity analysis and shown in Fig. 9 of ref. 8, as was the relative insensitivity of GEP to error in the isotopic composition of NEE and other measured variables.

Comparison with other methods for estimating GEP. There are other methods being developed for estimating GEP, including those based on measurements of carbonyl sulfide (OCS), whose uptake by plants has been taken as an index of photosynthetic CO₂ assimilation⁴⁰, and of sun-induced chlorophyll fluorescence (SIF)⁴¹, which is correlated with leaf photochemistry. Both OCS and SIF have been measured by others at or near this site for part of the period studied here (Extended Data Fig. 10). However, there are issues with both OCS and SIF that currently prevent these promising methods from estimating the seasonal pattern of GEP at the 5% uncertainty level needed to corroborate our present results. In the case of OCS, one issue is the occurrence of large unexplained emission of OCS at this site⁴⁰ (Extended Data Fig. 10). Another is that the uptake of OCS depends on stomatal conductance and carbonic anhydrase content, not on photosynthesis directly⁴⁰. In the case of SIF, the SIF–GEP relationship depends on the degree to which light is limiting photosynthesis (which depends on plant stress and carboxylation capacity)⁴² as well as on the proportion of sunlit and shaded leaves seen by the SIF sensor⁴¹, which varies with sun angle, diffuse light fraction, and canopy structure.

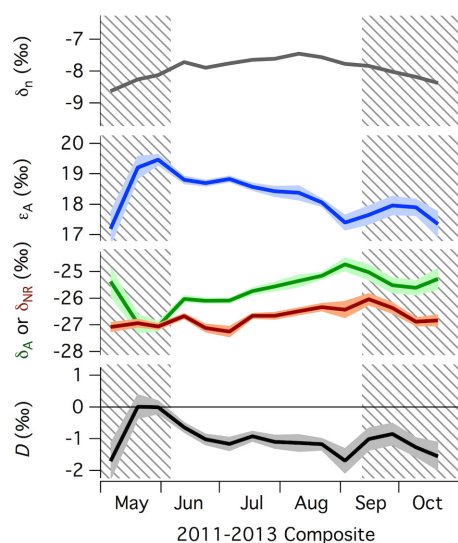
29. Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C. & Wofsy, S. C. Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Glob. Change Biol.* **2**, 169–182 (1996).
30. Yakir, D. & Wang, X. Fluxes of CO₂ and water between terrestrial vegetation and the atmosphere estimated from isotope measurements. *Nature* **380**, 515–517 (1996).
31. Bowling, D. R., Tans, P. P. & Monson, R. K. Partitioning net ecosystem carbon exchange with isotopic fluxes of CO₂. *Glob. Change Biol.* **7**, 127–145 (2001).
32. Ogée, J. *et al.* Partitioning net ecosystem carbon exchange into net assimilation and respiration using ¹³CO₂ measurements: a cost-effective sampling strategy. *Glob. Biogeochem. Cycles* **17**, 1070 (2003).
33. Lai, C.-T., Schauer, A. J., Owensby, C., Ham, J. M. & Ehleringer, J. R. Isotopic air sampling in a tallgrass prairie to partition net ecosystem CO₂ exchange. *J. Geophys. Res.* **108**, 4566 (2003).
34. Knoch, A. & Buchmann, N. Partitioning the net CO₂ flux of a deciduous forest into respiration and assimilation using stable carbon isotopes. *Global Biogeochem. Cycles* **19**, GB4008 (2005).
35. Zhang, J., Griffis, T. J. & Baker, J. M. Using continuous stable isotope measurements to partition net ecosystem CO₂ exchange. *Plant Cell Environ.* **29**, 483–496 (2006).
36. Zobitz, J. M., Burns, S. P., Reichstein, M. & Bowling, D. R. Partitioning net ecosystem carbon exchange and the carbon isotopic disequilibrium in a subalpine forest. *Glob. Change Biol.* **14**, 1785–1800 (2008).

37. Billmark, K. A. & Griffis, T. J. in *Phenology of Ecosystem Processes* 143–166 (Springer, 2009).
38. Fassbinder, J. J., Griffis, T. J. & Baker, J. M. Evaluation of carbon isotope flux partitioning theory under simplified and controlled environmental conditions. *Agric. For. Meteorol.* **153**, 154–164 (2012).
39. Ogee, J. *et al.* Partitioning net ecosystem carbon exchange into net assimilation and respiration with canopy-scale isotopic measurements: an error propagation analysis with $^{13}\text{CO}_2$ and CO^{18}O data. *Global Biogeochem. Cycles* **18**, GB2019 (2004).
40. Commane, R. *et al.* Seasonal fluxes of carbonyl sulfide in a mid-latitude forest. *Proc. Natl. Acad. Sci. USA* **112**, 14162–14167 (2015).
41. Yang, X. *et al.* Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. *Geophys. Res. Lett.* **42**, <http://dx.doi.org/10.1002/2015GL063201> (2015).
42. van der Tol, C., Berry, J. A. & Campbell, P. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. *J. Geophys. Res. Biogeosci.* **119**, <http://dx.doi.org/10.1002/2014JG002713> (2014).
43. Gatz, D. F. & Smith, L. The standard error of a weighted mean concentration—I. Bootstrapping vs other methods. *Atmos. Environ.* **29**, 1185–1193 (1995).



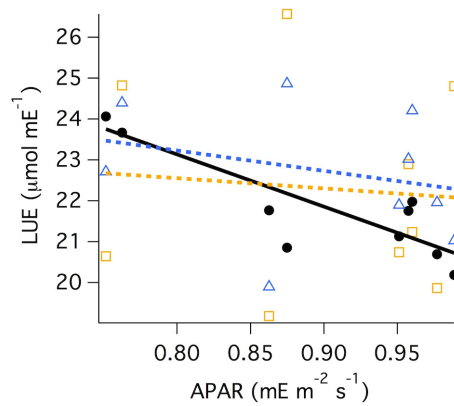
Extended Data Figure 1 | Accuracy and precision of long-term isotopic measurements. Repeated quantum cascade laser spectrometer measurements (dots, each representing a measurement integrated over 100 s) of the isotopic compositions of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ in a single known reference cylinder, but measured as if it were an unknown value interspersed among the three years of routine atmospheric measurements. Known reference cylinder values are indicated by the solid grey lines, with

95% confidence intervals indicated by the grey regions. Except for a period in September 2011 (between the vertical orange lines) when an inferior instrument thermal regulation scheme was tested, the precision of the spectrometer's rapid, *in situ* isotope measurements is seen to be better than that obtained for the reference cylinder by laboratory-based isotope ratio mass spectrometry⁷.



Extended Data Figure 2 | Composite seasonal cycles of isotopic compositions, isotopic discrimination and isotopic disequilibrium.

Shown are: the isotopic composition of CO_2 in the canopy airspace, δ_n ; the apparent fractionation by net photosynthetic assimilation (also called discrimination), ϵ_A ; the isotopic signatures of net photosynthetic assimilation, δ_A , and non-foliar respiration, δ_{NR} ; and the isotopic disequilibrium, $D = \delta_{NR} - \delta_A$. Dark lines connect flux-weighted means over all daylight hours for each 12-day bin, except in the case of δ_{NR} , where the lines connect simple means over all night-time hours for each bin (because δ_{NR} is derived from night-time Keeling plots rather than daytime flux measurements). Light shaded bands show standard errors in the flux-weighted means, calculated according to the ratio variance approximation recommended in ref. 43 (or just standard errors in the means for δ_{NR}), and based on variability within each bin ($64 \leq n \leq 431$ for daylight bins, and $16 \leq n \leq 33$ for δ_{NR}). Hatched areas indicate periods of leaf expansion and abscission.



Isotopic Partitioning

$$r^2 = 0.75; p = 0.003$$

$$\text{intercept} = 33.4 \pm 2.5; \text{slope} = -12.8 \pm 2.8$$

Partitioning from Ref. 13

$$r^2 = 0.07; p = 0.48$$

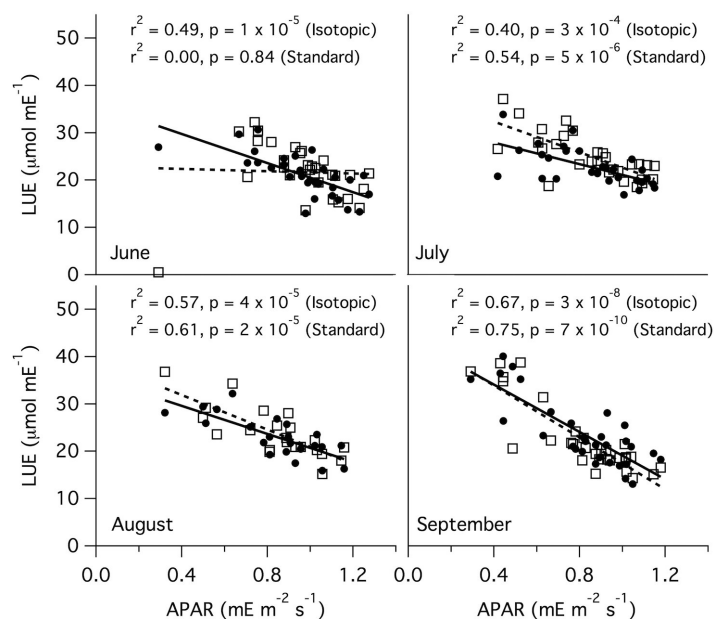
$$\text{intercept} = 27.2 \pm 6.0; \text{slope} = -5.0 \pm 6.6$$

Partitioning from Ref. 14

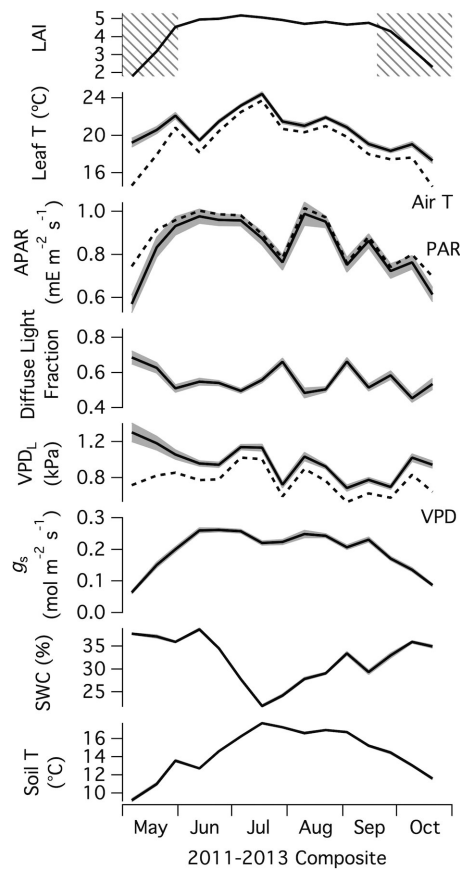
$$r^2 = 0.01; p = 0.82$$

$$\text{intercept} = 24.5 \pm 9.6; \text{slope} = -2.5 \pm 10.7$$

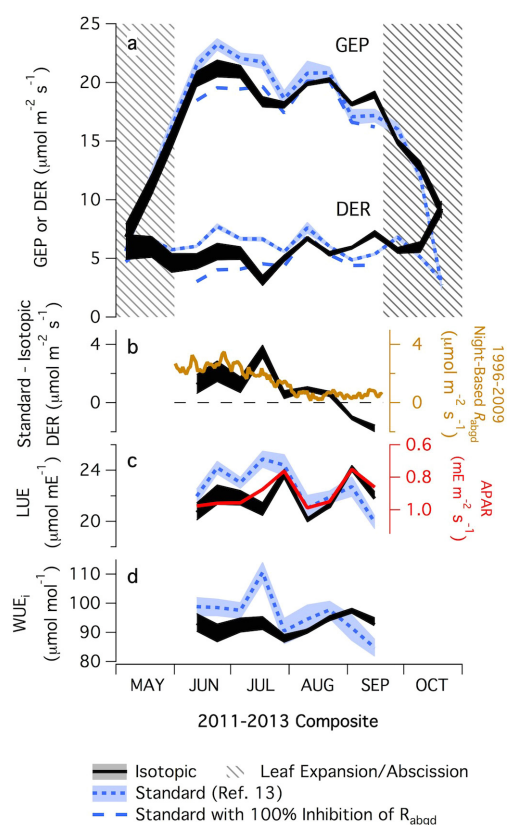
Extended Data Figure 3 | Relationships of LUE to APAR, from our isotopic partitioning and from both standard methods. Scatterplot of the LUE and APAR data from Fig. 2c (solid black circles), along with ordinary least-squares linear fits (black lines), for the period from full leaf expansion to the onset of senescence. These results are from partitioning based on isotopes. Also shown are results from the standard method of ref. 13 (hollow blue triangles), and from the partitioning method of ref. 14 (hollow yellow squares).



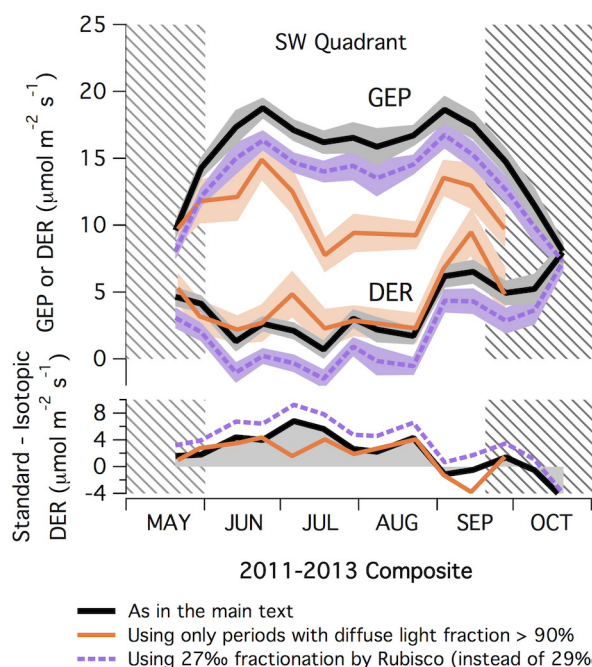
Extended Data Figure 4 | Relationships of LUE to APAR within each month. Daily LUE is plotted against daily APAR, averaged by day of year across all three years, on the basis of isotopic partitioning (solid circles) and the standard method of ref. 13 (hollow squares), and plotted separately for June, July, August and September. Also shown are linear (ordinary least-squares) fits for the isotopic (solid line) and standard (dotted line) partitioning methods.



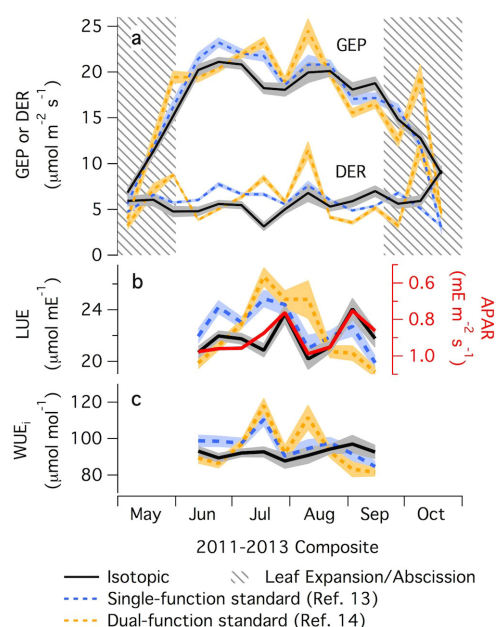
Extended Data Figure 5 | Composite seasonal cycles of environmental variables. Shown are leaf area index (LAI), leaf temperature (Leaf T), APAR, diffuse light fraction, leaf–air water vapour pressure difference (VPD_L), canopy stomatal conductance (g_s), volumetric soil water content (SWC), and soil temperature at 10 cm depth (Soil T), averaged across the three years, 2011–2013. Lines connect means over all daylight hours within each 12-day bin, and grey bands show standard errors in the means, calculated from variability within each bin ($64 \leq n \leq 431$). Air temperature (Air T), PAR, and the atmospheric water vapour pressure deficit (VPD) are also shown, as dotted lines. Hatched areas indicate leaf expansion and abscission.



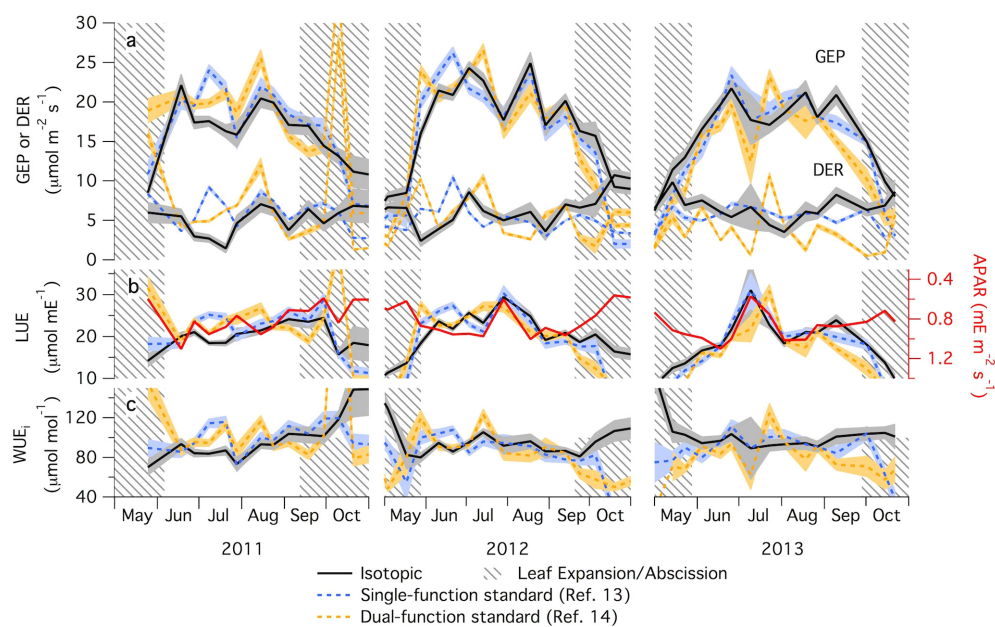
Extended Data Figure 6 | Effect of varying the prescribed rate of foliar respiration on the seasonal patterns of GEP and DER. As for Fig. 2, but with the black lines thickened to show the range of GEP and DER values that result from prescribing between 0% and 100% inhibition of leaf respiration by light. The grey standard error bands in Fig. 2 have been removed here for clarity. Hatched areas indicate leaf expansion and abscission. **a**, GEP and DER. **b**, Discrepancy between standard and isotopic partitioning (black line), with the gold line showing the 1996–2009 mean seasonal pattern of aboveground respiration (R_{abgd}) estimated from soil chambers and night-time NEE¹⁹. **c**, Light-use efficiency (LUE; isotopic and standard partitioning), with absorbed photosynthetically active radiation (APAR) inverted in red. **d**, Intrinsic water-use efficiency (WUE_i).



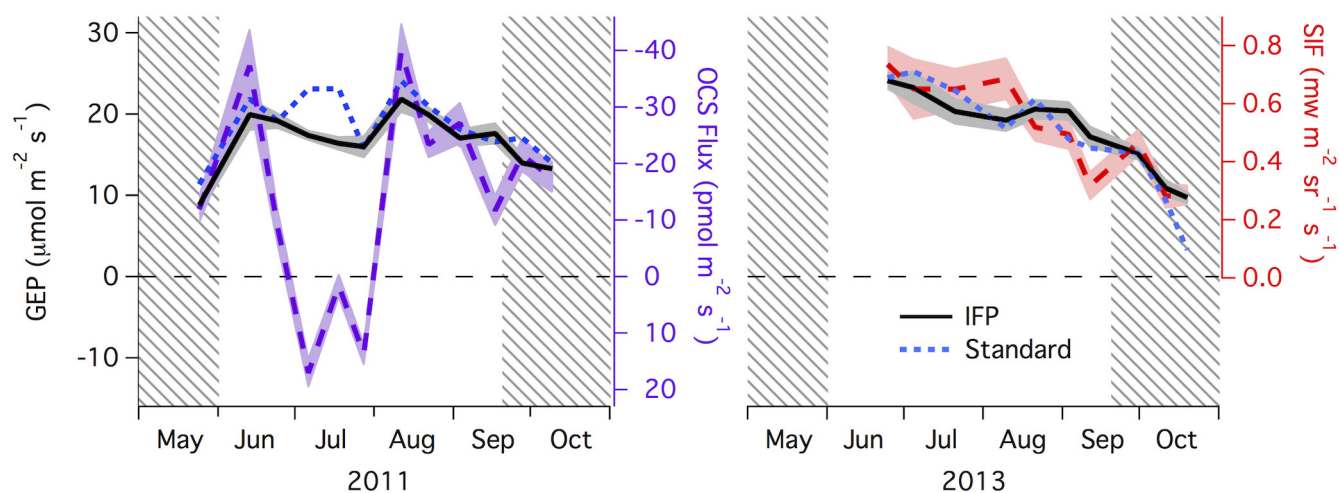
Extended Data Figure 7 | Sensitivity of the seasonal cycles of GEP and DER to change in the isotopic fractionation by the photosynthetic enzyme Rubisco, and to restriction to diffuse light conditions. This figure compares the composite seasonal cycles (across the three years, 2011–2013) of GEP and DER obtained from three variations of the IFP method (restricted to the southwest quadrant to reduce spurious discrepancies caused by differences in the flux tower sampling footprint when subsampling for diffuse light fraction). Top panel: GEP and DER from IFP. Bottom panel: the discrepancy between values of DER obtained from standard partitioning (based on night-time NEE and temperature), and values obtained from isotopic partitioning. The IFP variations shown are: as described in the text (solid black lines); restricted to periods with diffuse light fractions greater than 90% (solid orange lines); and using 27‰ instead of 29‰ for the isotopic fractionation by Rubisco-catalysed fixation of CO_2 (dotted purple lines). The lines connect the means (which are from all daylight hours) for each 12-day bin. The light shaded bands around each line in the top panel show the standard error of the mean, calculated from the variability within each bin ($25 \leq n \leq 130$). Hatched areas indicate periods of leaf expansion and abscission.



Extended Data Figure 8 | Composite seasonal cycles, from isotopic partitioning and from both standard partitioning methods. Shown are results from isotopic partitioning (solid black); from standard partitioning based on night-time NEE and temperature (dotted blue); and from standard partitioning incorporating a photosynthetic function of light (dotted yellow). **a**, GEP and DER. **b**, LUE, with APAR inverted in red. **c**, WUE_i . Lines connect means over all daylight hours for each 12-day bin; pale bands show standard errors of the means calculated from variability within each bin ($64 \leq n \leq 431$).



Extended Data Figure 9 | Seasonal cycles from isotopic partitioning and from both standard partitioning methods, for individual years. As for Extended Data Fig. 8, but showing the individual years separately ($8 \leq n \leq 204$).



Extended Data Figure 10 | Comparison of GEP values obtained from isotopic partitioning with preliminary estimates based on measurements of carbonyl sulfide and solar-induced fluorescence. Seasonal patterns of GEP from IFP (solid black) and from the standard method of ref. 13 (dotted blue) are compared with those of the OCS flux in 2011 (dashed purple, on an inverted scale) and the SIF signal in

2013 (dashed red). Lines connect means for each 12-day bin, and pale bands show standard errors of the means calculated from variability within each bin ($10 \leq n \leq 209$). The analysis included only data points for which simultaneous GEP and OCS, or GEP and SIF, measurements were available. The OCS data were provided by Commane *et al.*⁴⁰, and the SIF data by Yang *et al.*⁴¹.

Interdisciplinary research has consistently lower funding success

Lindell Bromham¹, Russell Dinnage¹ & Xia Hua¹

Interdisciplinary research is widely considered a hothouse for innovation, and the only plausible approach to complex problems such as climate change^{1,2}. One barrier to interdisciplinary research is the widespread perception that interdisciplinary projects are less likely to be funded than those with a narrower focus^{3,4}. However, this commonly held belief has been difficult to evaluate objectively, partly because of lack of a comparable, quantitative measure of degree of interdisciplinarity that can be applied to funding application data¹. Here we compare the degree to which research proposals span disparate fields by using a biodiversity metric that captures the relative representation of different fields (balance) and their degree of difference (disparity). The Australian Research Council's Discovery Programme provides an ideal test case, because a single annual nationwide competitive grants scheme covers fundamental research in all disciplines, including arts, humanities and sciences. Using data on all 18,476 proposals submitted to the scheme over 5 consecutive years, including successful and unsuccessful applications, we show that the greater the degree of interdisciplinarity, the lower the probability of being funded. The negative impact of interdisciplinarity is significant even when number of collaborators, primary research field and type of institution are taken into account. This is the first broad-scale quantitative assessment of success rates of interdisciplinary research proposals. The interdisciplinary distance metric allows efficient evaluation of trends in research funding, and could be used to identify proposals that require assessment strategies appropriate to interdisciplinary research⁵.

The 5th Annual Meeting of the Global Research Council in New Delhi in May 2016 focused on interdisciplinarity as one of its main topics of concern, reflecting increasing interest in research that breaks free of traditional discipline boundaries, and the growing concern that interdisciplinary research is not adequately supported under current funding structures. Funding agencies play a key role in shaping interdisciplinary research³, with both positive influence, such as dedicated programmes for interdisciplinary projects, and negative impacts, as perceived biases can discourage submission of interdisciplinary proposals to open funding calls. This leads to the 'paradox of interdisciplinarity': interdisciplinary research is often encouraged at policy level but poorly rewarded by funding instruments⁴. There is a clear need to test the widely held belief that interdisciplinary proposals fare poorly in competitive funding rounds: confirmation could prompt examination of evaluation strategies for interdisciplinary projects, while rejection of this claim might encourage more interdisciplinary proposals.

Critical to evaluation of current practice is the ability to compare levels of interdisciplinarity of research projects to track trends, evaluate outputs and compare success rates^{6,7}. Measures of interdisciplinarity have typically relied on textual references, detecting use of words such as 'interdisciplinarity'⁸, or bibliometric analysis, tracking patterns of author affiliation⁶ or citations within publications^{9–14}. But these approaches are limited in use for evaluating funding applications. Interpretation of the terms 'multidisciplinary', 'cross-disciplinary', 'interdisciplinary'

and 'transdisciplinary' vary widely¹¹, and researchers will differ in their inclination to label their research as 'interdisciplinary'², particularly if they perceive that identification as interdisciplinary influences funding outcomes¹⁵. Because bibliometric analyses are primarily applied to publications¹¹, they may be of limited applicability in assessing funding proposals, where the outputs have not yet been published, and citations may not be in an analysable format. The lack of clear definitions and objective analyses is an impediment to evaluating the relative success of interdisciplinary proposals¹. Conflicting findings have been reported using a range of approaches⁴, and most studies of funding success of interdisciplinary research have selected only a sample of proposals for evaluation^{15,16}. What is needed is a measure of the degree to which a proposal spans many different disciplines, independent of use of words such as 'interdisciplinarity' and without relying on cited publication data.

Although no single metric will capture all salient aspects of interdisciplinarity, developing a simple measure of the disciplinary spread of research proposals does provide a tractable way to compare the relative success of proposals having a narrow disciplinary focus with those with a broader research programme¹⁷. To this end, we use information supplied on funding applications to score each proposal on the disparity and balance of the component disciplines¹². We base our analysis on methods established in evolutionary biology to account for relatedness between biological lineages, but instead of using an evolutionary tree (phylogeny), we use a hierarchical classification of research fields. This metric can be applied to any funding scheme where multiple discipline categories can be selected by applicants or identified from proposal documents⁷.

We calculated this interdisciplinary distance (IDD) metric for all proposals submitted to the Australian Research Council Discovery Programme between 2010 and 2014 (Supplementary Table 1). This national competitive grants scheme funds fundamental research in all academic fields, receiving approximately 3,500 proposals in each annual funding call, with success rates being around 15–20% of proposals (Extended Data Table 1). Our analysis is unique in including all submitted proposals, both successful and unsuccessful, whereas most analyses are restricted to the published lists of funded proposals^{7,18} or to samples of case studies³.

Every application must nominate at least one of a defined set of 1,238 Field of Research codes, assigning a percentage weighting to each code selected. Field of Research codes are grouped into related disciplines: for example, the Division '06 Biological Sciences' contains nine groups, including '0603 Evolutionary Biology', which contains 12 Fields including '060309 Phylogeny and Comparative Analysis' (Supplementary Fig. 1). Because we wish to capture the disciplinary breadth of proposals, we need to measure not only the number and relative representation of disciplines selected, but also how disparate those research fields are. For example, we want to score a project that involves collaboration between biologists and artists as more interdisciplinary than one between biochemists and geneticists. Just as many biodiversity metrics use a phylogeny to measure disparity of species, research

¹Research School of Biology, Australian National University, 116 Daley Road, Canberra 0200, Australia.

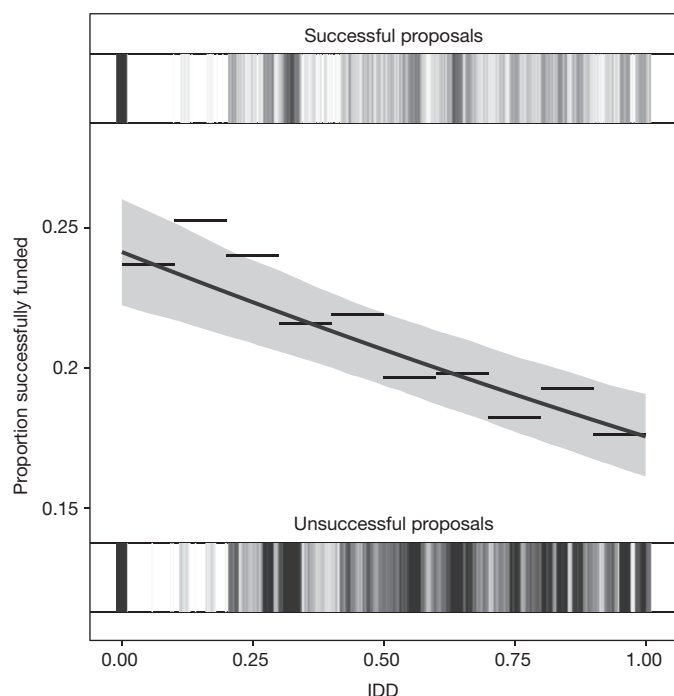


Figure 1 | Relationship between funding success and IDD score.

The central black line is the regression line (success = $\text{logit}^{-1}[-1.35506 - 0.40268 \times \text{IDD}]$); grey area represents the confidence intervals (see Supplementary Information for details). The horizontal lines around the regression line represent the mean success rate for each of 10 'bins' of IDD values of width 0.1. Each horizontal line's y value represents the mean success rate of proposals whose IDD values fall within a 'bin' defined by the ends of the horizontal line. The bars at the top and bottom of the figure indicate the number of proposals for each IDD score, with darker lines corresponding to more proposals. Fitted line based on a generalized linear mixed model as described in the Supplementary Information ($z = -6.789$; $P = 1.13 \times 10^{-11}$; $n = 18,476$).

fields can be arranged as a dendrogram, where fields of enquiry that are more similar are more closely connected to each other than more distant fields. Our approach is not restricted to funding programmes that use hierarchically structured research codes, but can be applied to any research field identifiers by using patterns of co-occurrence to define clusters of similar fields⁷.

We use the phylogenetic species evenness¹⁹ metric to measure IDD. This metric was designed to compare biodiversity between samples (for example, between conservation areas), incorporating both evenness of species representation in the biota and relatedness between species²⁰. IDD reflects both the relative contribution of disciplines within a proposal (balance) and scores collaborations between distantly related fields more highly than those between more closely related disciplines (disparity)²¹. The metric is standardized so it falls between 0 (single disciplinary) and 1 (maximum disparity with even representation), allowing direct comparison between proposals (Extended Data Fig. 1). Patterns of co-occurrence of field identifiers (for example, research codes or key words) can be used to generate a hierarchy of discipline relationships (see Supplementary Information). Basing this hierarchy on observed patterns of collaboration would rank proposals with respect to the relative novelty of the disciplinary combinations proposed²².

The Australian Research Council provided de-identified data on all applications to the Discovery Programme for five annual funding rounds. We used a generalized linear mixed model to ask whether the IDD score of grant proposals is associated with funding success. We included as variables in the analysis the year of application, number of Field of Research codes selected per proposal, number of named chief investigators and institution (grouped into higher education networks:

Extended Data Table 2). The response variable was a binary vector with two states: recommended for funding (1) or not recommended for funding (0). We provide details of the analysis and results in the Supplementary Information.

We find that IDD is consistently negatively correlated with funding success (slope = -0.40 , $P = 1.1 \times 10^{-11}$; Fig. 1), independent of year of application, number of research codes selected and primary research field. Nearly all research fields have reduced funding success with increasing interdisciplinarity (Fig. 2 and Extended Data Table 3). If the association between IDD and funding success was largely a matter of averaging success rates of the component disciplines—so that less successful fields benefit from collaboration with more successful fields but more successful fields have their success rates reduced through collaboration—then we would expect many points both above and below $y = 0$ in Fig. 2. However, most fields have negative values of y , suggesting that proposals with high IDD are expected to have lower success rates than those with low IDD in most research divisions.

We conducted additional analyses using metrics that reflected only variety (number of codes) or balance (evenness) of disciplines (details in Supplementary Information), which demonstrated that it is both disparity and balance between disciplines that influence chance of funding success, justifying the use of the IDD metric which captures both of these aspects of interdisciplinarity. We also searched for relevant keywords in proposal titles and summaries, such as 'interdisciplinary', 'multidisciplinary', 'transdisciplinary' and 'cross-disciplinary'. We found that proposals with these keywords also had higher IDD measures and lower success rates, demonstrating that the metric-based approach echoes text-based analysis, yet with greater power to detect differences in funding rates (see Supplementary Information).

IDD reflects the interdisciplinarity of the project, not the inclusion of practitioners from different disciplines, because a single researcher can devise a project that spans different disciplinary traditions¹¹. To test the influence of collaboration, we added the number of named chief investigators to the analysis. We find that proposals with more chief investigators have slightly higher success rates (slope = 0.03 , $P = 0.003$), across all research fields, independently of any link between number of collaborators and interdisciplinarity (Supplementary Table 2). Although the relationship between number of chief investigators and IDD is positive, the effect is small (Spearman's $\rho = 0.09$), suggesting that the number of participants is not strongly associated with the interdisciplinarity of the project proposal.

Because of the perceived negative association between funding success and interdisciplinarity, interdisciplinary projects are often regarded as high-risk proposals⁴. Do institutions with higher rates of funding success submit more interdisciplinary proposals (using higher success rates to support risky proposals) or fewer (because higher success rates arise from narrowly focused research)? We find that overall funding success rates varied between institutions, with significantly higher funding success rates in leading research-intensive universities (Extended Data Table 2)¹⁸. Differences in IDD between institutions were very small ($R^2 = 0.001$) and the negative relationship between IDD and success rate was significant when institution was taken into account (slope = -0.39 , $P = 7.6 \times 10^{-11}$; Supplementary Table 2). This suggests that the negative relationship between interdisciplinarity and funding success is not due to institutions with high funding success submitting more narrowly focused proposals (Extended Data Fig. 2).

Why do interdisciplinarity proposals have lower funding success rates? It is widely believed that grant evaluation processes are biased against interdisciplinary projects, because proposals may be assigned to a panel or reviewers who are ill-equipped to evaluate all parts of the project²³, while more narrowly focused proposals may be better matched to assessor expertise⁴. Proposals that fit within a well-defined discipline may be more easily explained and justified, whereas the novelty of combinations of different perspectives may be more difficult to explain²² or result in less-focused proposals²⁴.

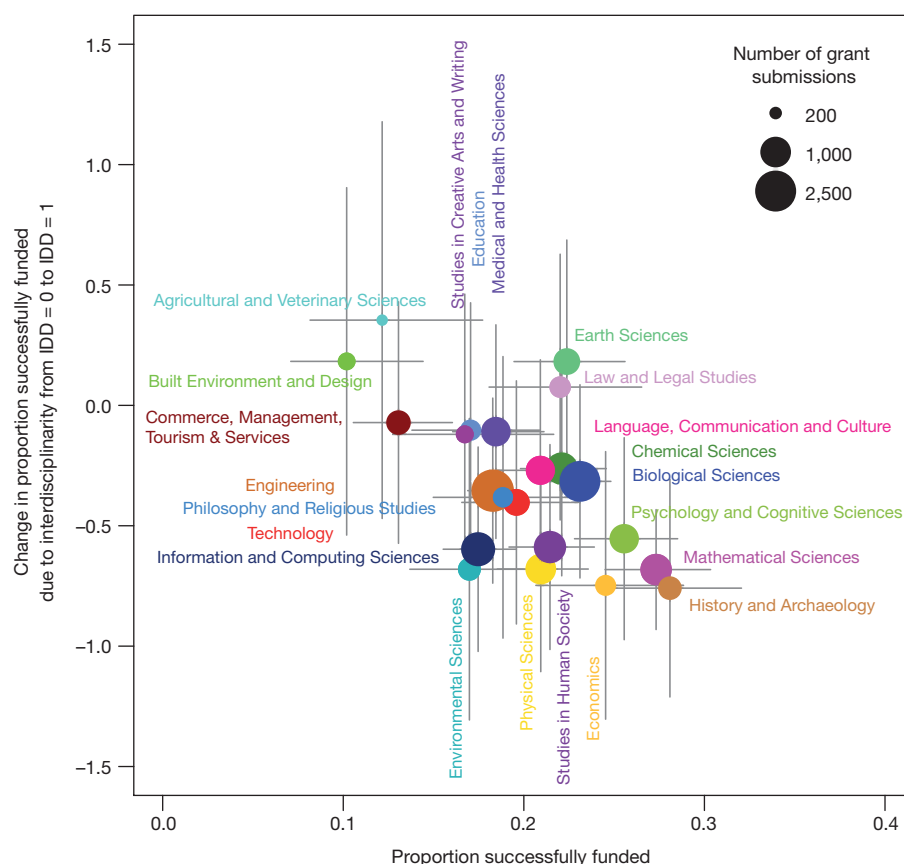


Figure 2 | Relationship between interdisciplinarity and funding success by research division. The x axis gives the success rate of a research division as the proportion of successfully funded proposals in that research division. The error bars along the x axis show the confidence interval of success rate approximated by Wilson interval. The number of proposals in each research division is given in Extended Data Table 1. The y axis gives the average predicted difference in logit success rates between proposals with maximum interdisciplinarity ($IDD = 1$) compared with proposals

with a single primary Field of Research code ($IDD = 0$) for each division, so that $y = -0.5$ indicates that the logit success rates of proposals with $IDD = 1$ is 0.5 lower than the logit success rate of proposals with $IDD = 0$. The standard error of the predicted difference (the error bar along the y axis) is the square root of the sum of the squared standard error of the IDD coefficient and that of the interaction coefficient. The average difference of each research division and its error bar are predicted by the generalized linear mixed model in Extended Data Table 3.

While interdisciplinary research can have considerable benefits, it can also incur substantial costs, owing to the need to invest significant time in building collaborative relationships, developing a shared language and honing a common perspective from disparate viewpoints²⁵. The outputs of interdisciplinary projects may be fewer and of different kinds to projects with a narrower disciplinary focus^{26,27}. Research evaluation systems with a narrow range of measures of success—for example, number of primary research publications in peer-reviewed journals—may disadvantage interdisciplinary proposals where some key outputs are less easily measured, such as the establishment of collaborative networks or data-sharing agreements²³. While some interdisciplinary studies produce significant advances, the average quality of interdisciplinary proposals may not be the same as more narrowly focused research. Studies of the long-term scholarly impact of interdisciplinary research have had mixed results: whereas some suggest greater benefits, others find no support for higher impact of interdisciplinary research^{12–14}.

Whatever the cause of the correlation, our result confirms the long-held belief that interdisciplinary proposals have lower funding success rates, providing the basis for further investigation into the development and evaluation of interdisciplinary research. Although IDD does not capture all key aspects of interdisciplinary research, it does provide a tractable and adaptable way of comparing interdisciplinarity between proposals and tracking trends in application rates and funding success. IDD can be applied to any funding programme where research fields are identified. Relatedness between disciplines can be defined a priori (for example, Field of Research codes), through clustering analysis

of previous applications (see, for example, Extended Data Table 1), subjectively (based on experience) or by any other relevant means⁷. Such analyses will bring much needed clarity to determining whether interdisciplinary research programmes are being adequately supported under current funding models. In addition to enabling assessment of biases in success rates, the IDD metric could provide a way of identifying highly interdisciplinary proposals that might require special evaluation strategies, such as seeking reviewers who have experience in research spanning multiple fields.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 December 2015; accepted 11 May 2016.

- Rylance, R. Grant giving: global funders to focus on interdisciplinarity. *Nature* **525**, 313–315 (2015).
- Ledford, H. How to solve the world's biggest problems. *Nature* **525**, 308–311 (2015).
- Lyall, C., Bruce, A., Marsden, W. & Meagher, L. The role of funding agencies in creating interdisciplinary knowledge. *Sci. Public Policy* **40**, 62–71 (2013).
- Woelert, P. & Millar, V. The “paradox of interdisciplinarity” in Australian research governance. *High. Educ.* **66**, 755–767 (2013).
- Committee on Facilitating Interdisciplinary Research. *Facilitating Interdisciplinary Research* (National Academy of Sciences, National Academy of Engineering, Institute of Medicine, 2004).
- Langfeldt, L. The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. *Res. Eval.* **15**, 31–41 (2006).
- Nichols, L. G. A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics* **100**, 741–754 (2014).

8. Van Noorden, R. Interdisciplinary research by the numbers. *Nature* **525**, 306–307 (2015).
9. Porter, A. & Rafols, I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* **81**, 719–745 (2009).
10. Porter, A. L., Roessner, J. D., Cohen, A. S. & Perreault, M. Interdisciplinary research: meaning, metrics and nurture. *Res. Eval.* **15**, 187–195 (2006).
11. Wagner, C. S. *et al.* Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. *J. Informetrics* **5**, 14–26 (2011).
12. Yegros-Yegros, A., Rafols, I. & D'Este, P. Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity. *PLoS ONE* **10**, e0135095 (2015).
13. Wang, J., Thijs, B. & Glänzel, W. Interdisciplinarity and impact: distinct effects of variety, balance, and disparity. *PLoS ONE* **10**, e0127298 (2015).
14. Shi, X., Adamic, L. A., Tseng, B. L. & Clarkson, G. S. The impact of boundary spanning scholarly publications and patents. *PLoS ONE* **4**, e6547 (2009).
15. Huutoniemi, K., Klein, J. T., Bruun, H. & Hukkinen, J. Analyzing interdisciplinarity: typology and indicators. *Res. Policy* **39**, 79–88 (2010).
16. Bruun, H., Hukkinen, J., Huutoniemi, K. & Klein, J. T. *Promoting Interdisciplinary Research: The Case of the Academy of Finland* (The Academy of Finland, 2005).
17. Bammer, G. *Strengthening Interdisciplinary Research: What It Is, What It Does, How It Does It and How It Is Supported* (Australian Council of Learned Academies, 2012).
18. Ma, A., Mondragón, R. J. & Latora, V. Anatomy of funded research in science. *Proc. Natl Acad. Sci. USA* **112**, 14760–14765 (2015).
19. Helmus, M. R., Bland, T. J., Williams, C. K. & Ives, A. R. Phylogenetic measures of biodiversity. *Am. Nat.* **169**, E68–E83 (2007).
20. Cadotte, M. W. *et al.* Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecol. Lett.* **13**, 96–105 (2010).
21. Stirling, A. A general framework for analysing diversity in science, technology and society. *J. R. Soc. Interface* **4**, 707–719 (2007).
22. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
23. Porter, A. L., Garner, J. & Crowl, T. Research coordination networks: evidence of the relationship between funded interdisciplinary networking and scholarly impact. *Bioscience* **62**, 282–288 (2012).
24. Boix Mansilla, V., Feller, I. & Gardner, H. Quality assessment in interdisciplinary research and education. *Res. Eval.* **15**, 69–74 (2006).
25. Haythornthwaite, C., Lunsford, K. J., Bowker, G. C. & Bruce, B. C. in *New Infrastructures for Science Knowledge Production* (ed. Hine, C.) 143–166 (Idea Group, 2006).
26. Laudel, G. Conclave in the Tower of Babel: how peers review interdisciplinary research proposals. *Res. Eval.* **15**, 57–68 (2006).
27. Goring, S. J. *et al.* Improving the culture of interdisciplinary collaboration in ecology by expanding measures of success. *Front. Ecol. Environ* **12**, 39–47 (2014).

Supplementary Information is available in the online version of the paper.

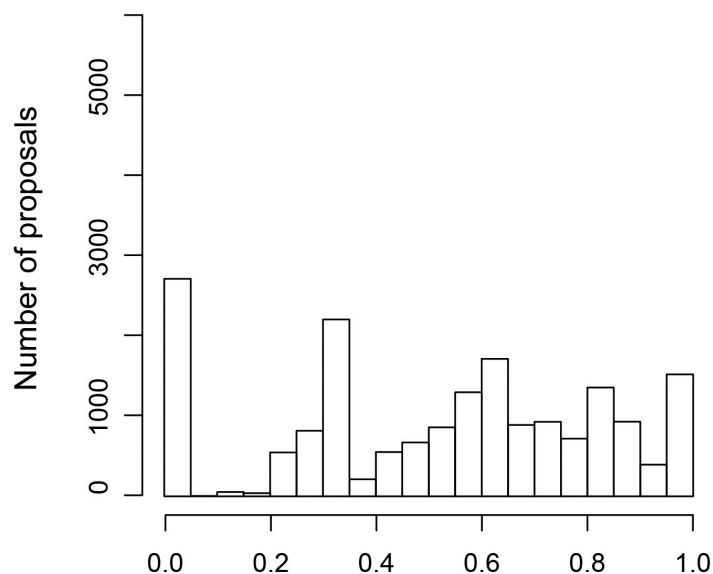
Acknowledgements We thank the Australian Research Council for providing de-identified application data for analysis, and for their commitment to transparency and improvement of research proposal assessment. We are grateful to A. Byrne for his feedback and encouragement. We also thank M. Jennions for feedback, and G. Bammer, J. Bennett and the participants of the workshop on Interdisciplinary Research: Evaluating and Rewarding High-Quality Projects held at the University of New South Wales in August 2015.

Author Contributions All authors contributed equally to this work. L.B. conceived the project and wrote the paper; R.D. and X.H. designed, conducted and interpreted the analyses.

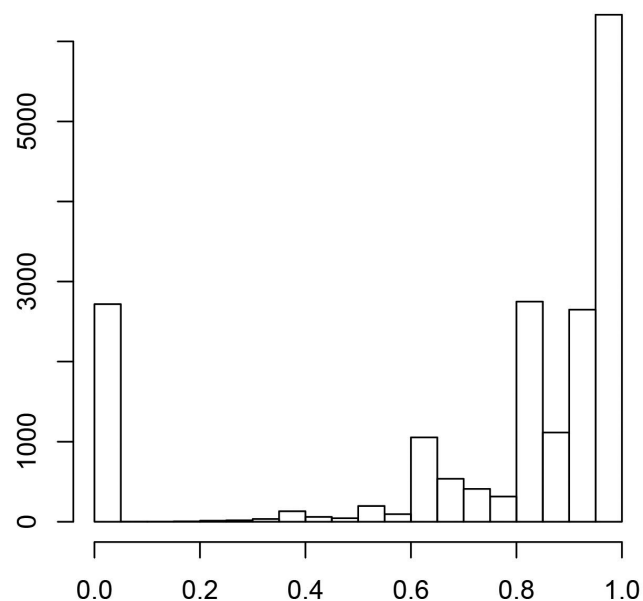
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.B. (lindell.bromham@anu.edu.au).

Reviewer Information *Nature* thanks L. Amaral, M. Helmus and the other anonymous reviewer(s) for their contribution to the peer review of this work.

A) Observed distribution of IDD



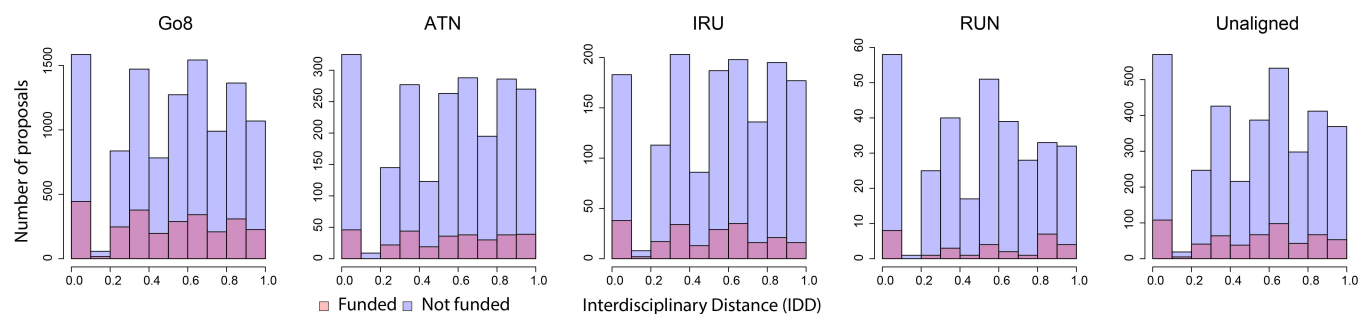
B) Null distribution of IDD



Interdisciplinary Distance (IDD)

Extended Data Figure 1 | Comparison of observed distribution of IDD scores to a null distribution. **a**, Distribution of IDD scores for 18,476 proposals to the Australian Research Council Discovery Programme, pooled over 5 years (2010–2014). **b**, Null distribution of IDD scores

generated by random sampling of Field of Research codes conditional on the observed frequencies of number of selected codes and percentage allocations.



Extended Data Figure 2 | Distribution of IDD scores by institutional networks. See Extended Data Table 2 for the membership of research networks. The research-intensive Group of Eight (Go8) universities submit

more proposals to the Australian Research Council Discovery Programme and have higher funding success rates, but the overall patterns of interdisciplinarity scores and success rates are similar across institutions.

Extended Data Table 1 | Summary of proposals submitted to Australian Research Council Discovery Programme between 2010 and 2014

Domain	Divisions	Number of Proposals	% Success	Median IDD	Median codes
01 Information	01 Mathematical Sciences	893	27	0.54	3
	08 Information And Computing Sciences	1351	17	0.54	2
02 Matter	02 Physical Sciences	1003	21	0.51	3
	03 Chemical Sciences	1172	22	0.63	3
	09 Engineering	2823	18	0.56	3
	10 Technology	582	20	0.80	3
03 Environment	04 Earth Sciences	706	22	0.54	3
	05 Environmental Sciences	418	17	0.75	3
04 Life	06 Biological Sciences	2458	23	0.56	3
	07 Agricultural And Veterinary Sciences	181	12	0.77	3
	11 Medical And Health Sciences	905	18	0.65	3
05 Interaction	12 Built Environment And Design	265	10	0.51	2
	14 Economics	424	25	0.48	3
	15 Commerce, Management, Tourism & Services	582	13	0.33	2
06 Mind	13 Education	434	17	0.56	2
	17 Psychology And Cognitive Sciences	896	26	0.32	2
	19 Studies In Creative Arts And Writing	269	17	0.64	2
07 Society	16 Studies In Human Society	1161	21	0.56	3
	18 Law And Legal Studies	368	22	0.32	3
	20 Language, Communication And Culture	722	21	0.55	3
	21 History And Archaeology	534	28	0.33	2
	22 Philosophy And Religious Studies	329	19	0.50	3

The number of proposals with a primary identification to each Division in 5 years of pooled applications to the Australian Research Council Discovery Programme, with the percentage recommended for funding (% success), median interdisciplinary distance (IDD) and median number of six-digit FOR codes selected per application (Median codes). Divisions (first two digits of the FOR codes) have been clustered into Domains, as described in the Supplementary Information. Note that medical research is predominantly funded through a different scheme, as is research in collaboration with industry partners.

Extended Data Table 2 | Institutional networks

Network		Institutions	Proposals	Success	IDD
Group of Eight	Go8	University of Melbourne Australian National University University of Sydney University of Queensland University of Western Australia University of Adelaide Monash University University of New South Wales	10974	24%	0.56
Innovative Research Universities	IRU	Charles Darwin University James Cook University Murdoch University Flinders University LaTrobe University Griffith University.	1486	15%	0.57
Australian Technology Network	ATN	Queensland University of Technology University of Technology Sydney RMIT University University of South Australia Curtin University.	2181	14%	0.57
Regional Universities Network	RUN	Central Queensland University Federation University Australia Southern Cross University University of New England University of Southern Queensland University of the Sunshine Coast	324	10%	0.56
Unaligned			3476	17%	0.56

Australian higher education institutions grouped by network, with the number proposals submitted to the Australian Research Council Discovery Programme over 5 years (2010–2014), the average percentage success rate (percentages that were recommended for funding) and the median IDD score of all proposals.

Extended Data Table 3 | Effect size of interdisciplinarity on funding success in each division

Division	Coeff	Int'n	Mean Cohen's D					Variance Cohen's D					Wtd avg
			Y1	Y2	Y3	Y4	Y5	Y1	Y2	Y3	Y4	Y5	
Mathematical Sciences	0	0	-0.24	-0.21	-0.26	-0.03	-0.29	0.16	0.16	0.18	0.17	0.18	-0.20
Physical Sciences	-0.37***	0.00	0.02	-0.32	-0.81	0.03	-0.16	0.15	0.20	0.19	0.18	0.18	-0.23
Chemical Sciences	-0.25*	0.42	-0.10	0.12	-0.27	-0.15	0.11	0.14	0.16	0.16	0.16	0.18	-0.06
Earth Sciences	-0.25*	0.86*	-0.07	-0.27	-0.01	0.13	0.01	0.18	0.21	0.21	0.21	0.23	-0.04
Environmental Sciences	-0.48**	0.00	-0.22	-0.19	0.10	-0.34	-0.10	0.25	0.31	0.34	0.29	0.30	-0.16
Biological Sciences	-0.20*	0.37	-0.09	-0.17	-0.03	0.08	-0.16	0.09	0.11	0.11	0.12	0.11	-0.08
Agricultural & Veterinary Sciences	-1.02***	1.04	0.14	0.21	-0.24	-0.04	-0.26	0.75	0.63	0.44	0.55	0.45	-0.07
Information & Computing Sciences	-0.60***	0.08	-0.24	-0.01	-0.18	-0.39	-0.04	0.15	0.18	0.15	0.16	0.17	-0.18
Engineering	-0.50***	0.33	-0.23	-0.03	-0.11	-0.08	-0.13	0.10	0.11	0.11	0.11	0.11	-0.11
Technology	-0.35*	0.28	0.11	-0.10	-0.41	-0.05	-0.20	0.21	0.24	0.25	0.28	0.22	-0.12
Medical & Health Sciences	-0.50***	0.57	-0.01	0.15	-0.13	-0.41	0.04	0.16	0.17	0.22	0.22	0.24	-0.06
Built Environment & Design	-1.17***	0.86	0.16	0.37	0.03	-0.67	-0.42	0.37	0.45	0.74	0.73	0.39	-0.07
Education	-0.59***	0.58	0.25	-0.13	-0.34	0.05	-0.18	0.26	0.31	0.29	0.32	0.28	-0.07
Economics	-0.17	-0.07	-0.23	-0.07	-0.05	-0.26	-0.24	0.27	0.24	0.29	0.22	0.28	-0.17
Commerce, Management, Tourism & Services	-0.90***	0.61	-0.21	-0.27	0.19	-0.01	0.65	0.24	0.24	0.30	0.30	0.43	0.01
Studies In Human Society	-0.31**	0.09	-0.32	-0.05	-0.03	-0.10	-0.36	0.16	0.17	0.16	0.16	0.17	-0.17
Psychology & Cognitive Sciences	-0.15	0.13	-0.41	-0.06	0.06	-0.21	-0.35	0.17	0.17	0.17	0.18	0.18	-0.19
Law & Legal Studies	-0.26+	0.76	0.22	0.30	-0.57	-0.07	-0.15	0.27	0.28	0.28	0.29	0.35	-0.05
Studies In Creative Arts & Writing	-0.60***	0.56	-0.46	-0.34	0.70	-0.32	0.37	0.38	0.36	0.38	0.39	0.37	-0.01
Language, Communication & Culture	-0.35**	0.41	0.00	-0.18	-0.03	-0.13	-0.10	0.17	0.21	0.21	0.22	0.24	-0.08
History & Archaeology	-0.03	-0.08	-0.37	-0.09	-0.51	-0.18	-0.10	0.19	0.21	0.24	0.25	0.23	-0.25
Philosophy & Religious Studies	-0.48**	0.30	-0.11	-0.20	-0.11	-0.31	0.21	0.27	0.31	0.31	0.39	0.37	-0.11

The generalized linear mixed model used to predict success rate has IDD, division and their interaction as fixed variables and year as a random variable. 'Mathematical Sciences' is used as the reference category for other divisions, so its coefficient and interaction are zeros. Coefficients show differences in success rates of other divisions compared with Mathematical Sciences. Interaction shows differences in the effect of interdisciplinarity on success rates of other divisions compared with Mathematical Sciences. The coefficient of IDD = -0.68**. A likelihood ratio test suggests that including IDD as a fixed variable significantly increases model fit to the data ($\chi^2_{22} = 57.94$, $P = 4.5 \times 10^{-5}$). *** $P = 0$; ** $P = 0.001$; * $P = 0.01$; + $P = 0.05$. Coeff, coefficient; Int'n, interaction; Y, year; Wtd avg, weighted average.

Basal forebrain projections to the lateral habenula modulate aggression reward

Sam A. Golden^{1,2}, Mitra Heshmati^{1,2*}, Meghan Flanigan^{1,2*}, Daniel J. Christoffel³, Kevin Guise^{1,2}, Madeline L. Pfau^{1,2}, Hossein Aleyasin¹, Caroline Menard¹, Hongxing Zhang⁴, Georgia E. Hodes¹, Dana Bregman¹, Lena Khibnik¹, Jonathan Tai¹, Nicole Rebusi¹, Brian Krawitz^{1,2}, Dipesh Chaudhury⁴, Jessica J. Walsh³, Ming-Hu Han^{1,4}, Matt L. Shapiro¹ & Scott J. Russo¹

Maladaptive aggressive behaviour is associated with a number of neuropsychiatric disorders¹ and is thought to result partly from the inappropriate activation of brain reward systems in response to aggressive or violent social stimuli². Nuclei within the ventromedial hypothalamus^{3–5}, extended amygdala⁶ and limbic⁷ circuits are known to encode initiation of aggression; however, little is known about the neural mechanisms that directly modulate the motivational component of aggressive behaviour⁸. Here we established a mouse model to measure the valence of aggressive inter-male social interaction with a smaller subordinate intruder as reinforcement for the development of conditioned place preference (CPP). Aggressors develop a CPP, whereas non-aggressors develop a conditioned place aversion to the intruder-paired context. Furthermore, we identify a functional GABAergic projection from the basal forebrain (BF) to the lateral habenula (IHb) that bi-directionally controls the valence of aggressive interactions. Circuit-specific silencing of GABAergic BF–IHb terminals of aggressors with halorhodopsin (NpHR3.0) increases IHb neuronal firing and abolishes CPP to the intruder-paired context. Activation of GABAergic BF–IHb terminals of non-aggressors with channelrhodopsin (ChR2) decreases IHb neuronal firing and promotes CPP to the intruder-paired context. Finally, we show that altering inhibitory transmission at BF–IHb terminals does not control the initiation of aggressive behaviour. These results demonstrate that the BF–IHb circuit has a critical role in regulating the valence of inter-male aggressive behaviour and provide novel mechanistic insight into the neural circuits modulating aggression reward processing.

To study individual differences in aggression, we adapted the sensory contact model of social defeat for CD-1 mice^{9–11}, which exhibit a wide spectrum of aggressive behaviours. In this procedure (Fig. 1a), a sexually experienced adult male CD-1 mouse is presented with a series of novel 6–8-week-old subordinate male C57BL/6J intruder mice, who do not themselves exhibit any aggressive behaviours towards CD-1 mice (Extended Data Fig. 1a–i). This procedure identifies individual differences in antagonist aggressive behaviours without producing lasting stress-related behavioural phenotypes (Extended Data Table 1). Ethological analysis revealed that approximately 70% (310/448) of mice exhibited aggressive behaviour (termed aggressors (AGGs)) during at least one session, while approximately 30% (138/448) failed to initiate aggressive behaviour (termed non-aggressors (NONs)) at any time (Fig. 1b).

After repeated intruder interactions, AGGs have elevated serum testosterone (Fig. 1c) and decreased corticosterone (Fig. 1d) levels relative to NONs, suggesting that NONs may be less dominant and experience forced intruder interactions as more stressful. Analysis of several common metrics for aggression revealed normalized

distributions across AGGs that increased between screening sessions (Fig. 1e, f and Extended Data Fig. 2a–g). Importantly, the mean number of attack bouts (Extended Data Fig. 2f) and mean duration of attack bouts (Extended Data Fig. 2g) significantly correlate to mean attack latency. Therefore, attack latency provides a reliable index of aggression behaviours. Subsequently, we focused on AGGs that exhibited attack latencies within the most aggressive quartile of the sample distribution. These data confirm that outbred CD-1 mice exhibit a wide spectrum of aggressive behaviour and physiological responses to an intruder, leading us to hypothesize that there may be differences in the valence of intruder interactions among AGGs and NONs.

To assay the motivational state associated with intruder pairings, we developed an aggression-based CPP procedure. In this model, CD-1 mice are screened for aggression phenotype and then conditioned for CPP (Fig. 1g) by receiving novel C57BL/6J intruder-paired or intruder-unpaired sessions twice a day for three days. AGGs show a CPP for the intruder-paired context, while NONs show a conditioned place aversion (CPA) (Fig. 1h–j and Extended Data Fig. 3a–d). CPA in NONs does not appear to result from baseline differences in mood and anxiety or lack of interest in social targets (Extended Data Tables 1 and 2). However, we found that the valence of intruder interactions in AGGs and NONs is dependent upon intruder mice being freely moving and physically accessible during conditioning. Using a sensory CPP procedure in which the intruder mouse is placed in a protective cage within the intruder-paired context, both CPP and CPA are abolished (Fig. 1k–n and Extended Data Fig. 3e–h). These data demonstrate individual differences in the positive or negative valence of intruder interactions in AGGs versus NONs.

Clinical^{2,12} and preclinical⁸ studies have implicated BF structures, such as the nucleus accumbens (NAc), lateral septum and diagonal band nuclei (DBN), as potentially important brain regions controlling aggression-related behaviours. However, there has been limited functional evidence that the BF, or its projections, directly modulate the rewarding aspects of aggression. To define BF projections, we injected an adeno-associated virus (AAV) vector expressing enhanced yellow fluorescent protein (eYFP) under a neuronal-specific human synapsin (hSyn) promoter (AAV2-hSyn-eYFP) into the BF of CD-1 mice (Fig. 2a–c, top, and Extended Data Fig. 4a–c) targeted specifically to the more anterior septo-accumbal transition zone of the basal forebrain¹³ and observe a prominent axonal projection to the IHb (Fig. 2b, top).

To characterize BF–IHb projections further, we injected the IHb (Fig. 2a, b, bottom) with a retrograde monosynaptic glycoprotein-dead rabies virus (G-deleted-rabies-eGFP)¹⁴. Within the anterior BF that overlaps with our anterograde viral infection, we observed retrograde labelling in the septum (~45%), DBN (~35%) and the medial NAc shell (~15%) (Fig. 2c, d, bottom). Within retrogradely labelled BF slices, we performed *in situ* hybridization for GAD67, a marker of inhibitory

¹Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²Graduate Program in Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³Department of Psychiatry and Behavioral Sciences, Stanford University Medical Center, Palo Alto, California 94305, USA.

⁴Pharmacology and Systems Therapeutics and Institute for Systems Biomedicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

*These authors contributed equally to this work.

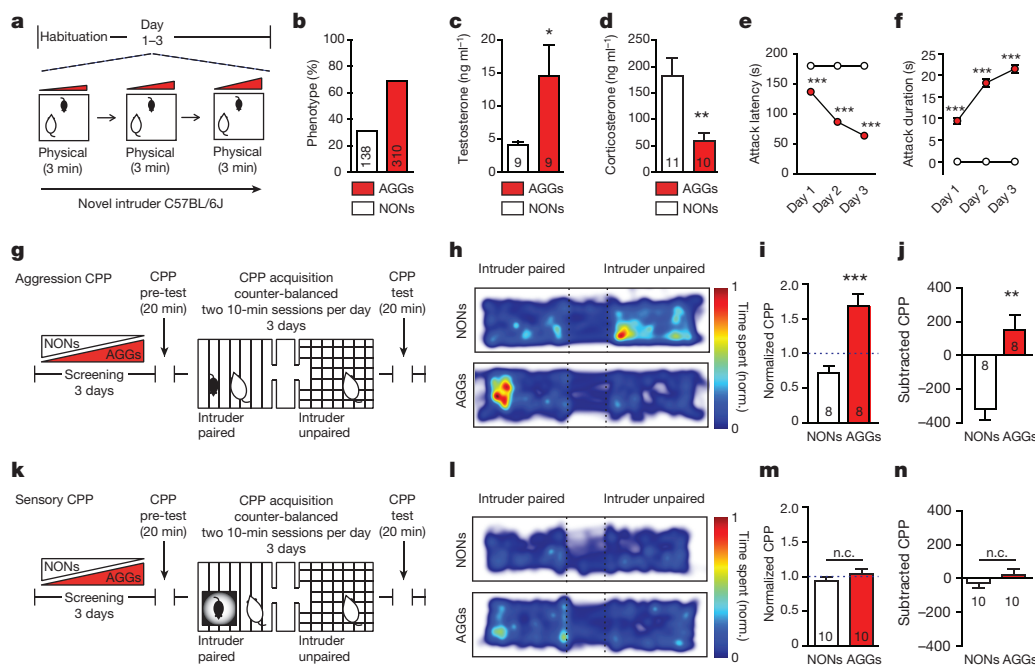


Figure 1 | Individual differences in aggression-related reward behaviour. **a**, Aggression screening: experimental schematic. **b**, Percentage of mice exhibiting aggressive (AGG) versus non-aggressive (NON) behaviours. **c**, **d**, Serum testosterone ($t_{16} = 2.23$, $*P < 0.05$; two-tailed unpaired t -test, $n = 9$ per group) (**c**) and corticosterone ($t_{19} = 3.231$, $**P < 0.01$; two-tailed unpaired t -test, $n = 10$ –11 per group) (**d**). **e**, **f**, Mean latency to attack (**e**) ($F_{2,1338} = 49.37$, two-way analysis of variance (ANOVA) $P < 0.001$; post-hoc test, $**P < 0.001$; $n = 138$ –310) and attack duration (**f**) ($F_{2,1338} = 22.35$, two-way ANOVA $P < 0.001$; post-hoc test, $**P < 0.001$; $n = 138$ –310). **g**, Aggression CPP schematic.

h, Representative heatmaps of aggression CPP. norm., normalized. **i**, **j**, Normalized ($t_{14} = 4.706$, $**P < 0.001$; two-tailed unpaired t -test, $n = 8$ per group) (**i**) and subtracted CPP score ($t_{14} = 4.013$, $**P < 0.01$; two-tailed unpaired t -test, $n = 8$ per group) (**j**). **k**, Sensory CPP schematic. **l**, Representative heatmaps of sensory CPP. **m**, **n**, Normalized (**m**) ($t_{18} = 1.023$, $P > 0.05$; two-tailed unpaired t -test, $n = 10$ per group) and subtracted (**n**) CPP score ($t_{18} = 0.961$, $P > 0.05$; two-tailed unpaired t -test, $n = 10$ per group). Summary data are represented as mean \pm standard error of the mean. n.c., no change. Experiments were conducted once; n indicates biological replicates.

GABAergic neurons, and observed colocalization within the septum (~75%), DBN (~80%) and medial NAc shell (100%) (Fig. 2e).

To identify whether BF and lHb neurons are differentially activated by intruder interactions in AGGs and NONs, we examined c-Fos immunoreactivity 1 h after the final intruder screening. AGGs exhibit elevated c-Fos immunoreactivity in the septo-accumbal transition zone of the BF relative to NONs (Fig. 2f, g). Within the eYFP-positive BF terminal fields in the medial lHb, NONs exhibit increased c-Fos immunoreactive nuclei relative to AGGs (Fig. 2f, g). This finding was corroborated by slice electrophysiology, in which NONs exhibit an increase in lHb firing rates compared with AGGs 1 h after an intruder interaction that returns to baseline by 7 days after intruder interaction (Fig. 2h, i). Together, these data show that lHb neurons are differentially regulated by intruder interactions, possibly through inhibitory BF inputs.

To determine the functional contribution of BF–lHb projections, we conducted optogenetic circuit-specific terminal photostimulation in combination with slice electrophysiology with channelrhodopsin (AAV2-hSyn-ChR2(H134R)-eYFP) or halorhodopsin (AAV2-hSyn-NpHR3.0-eYFP), identifying photostimulation parameters that produce robust transient lHb activation or inhibition without rebound neuronal firing. ChR2^{BF→lHb} terminal photostimulation with 40 Hz resulted in significantly decreased lHb firing rates (Fig. 2j, k), while NpHR3^{BF→lHb} (8 s on, 2 s off) terminal photostimulation resulted in a robust increase in postsynaptic lHb firing rates (Fig. 2l, m). Importantly, whole-cell recordings from lHb neurons during ChR2^{BF→lHb} terminal photostimulation showed a significant increase in inhibitory postsynaptic currents (IPSCs) that was completely blocked by the GABA_A receptor antagonist, gabazine (Extended Data Fig. 4d, e). Optically induced IPSCs exhibited a response delay of ~7 ms (Extended Data Fig. 4f), which is in line with previously published response delays for ChR2 at monosynaptic circuits. Similarly, anterograde tracing of BF terminals in the lHb revealed that they were colocalized with vesicular

GABA transporter (VGAT), but not vesicular glutamate transporter 1 (VGLUT1) (Extended Data Fig. 4g). To validate these findings within an intact system, we used multi-electrode recording of postsynaptic lHb firing rates in anaesthetized mice in combination with terminal photostimulation (Extended Data Fig. 5a). Results show that activation (40 Hz ChR2^{BF→lHb}), or inhibition (8 s on, 2 s off NpHR3^{BF→lHb}) of presynaptic BF terminals in the lHb resulted in decreased or increased lHb postsynaptic neuronal firing, respectively (Extended Data Fig. 5b–d). These functional *in vitro* and *in vivo* recordings of ChR2^{BF→lHb} and NpHR3^{BF→lHb} confirm inhibitory GABAergic control over circuit activity and demonstrate reliable temporal control of lHb firing rates by optogenetic tools for *in vivo* behavioural analysis.

To investigate the functional consequences of BF–lHb neuronal firing on aggression reward, we paired photostimulation of ChR2^{BF→lHb} and NpHR3^{BF→lHb} in AGGs and NONs during the CPP test (Fig. 3a, b). NON::ChR2^{BF→lHb} stimulation promoted CPP (Fig. 3c–e), mimicking responses observed in control AGGs. Conversely, AGG::NpHR3^{BF→lHb} stimulation induced CPA (Fig. 3f–h), mimicking responses observed in control NONs. Neither NON::NpHR3^{BF→lHb} or AGG::ChR2^{BF→lHb} stimulation significantly affected the expression of CPP or CPA. Viral expression (Extended Data Fig. 6a–f) and locomotor activity (Extended Data Fig. 6g–j) were not different between conditions. These data confirm that BF–lHb circuitry modulates the rewarding component of aggressive behaviour and is both necessary and sufficient for the expression of CPP in AGGs and CPA in NONs.

To determine if these circuit-specific effects could be recapitulated by direct lHb cell body manipulation, we injected the lHb with AAV2-hSyn-ChR2-eYFP or AAV-hSyn-NpHR3.0-eYFP (Extended Data Fig. 7a–d) and directly stimulated lHb cell bodies using previously established optogenetic parameters for lHb¹⁵. NON::NpHR3.0^{lHb} stimulation to decrease lHb firing promoted CPP to the intruder-paired side (Extended Data Fig. 7e–g), whereas AGG::ChR2^{lHb} stimulation to

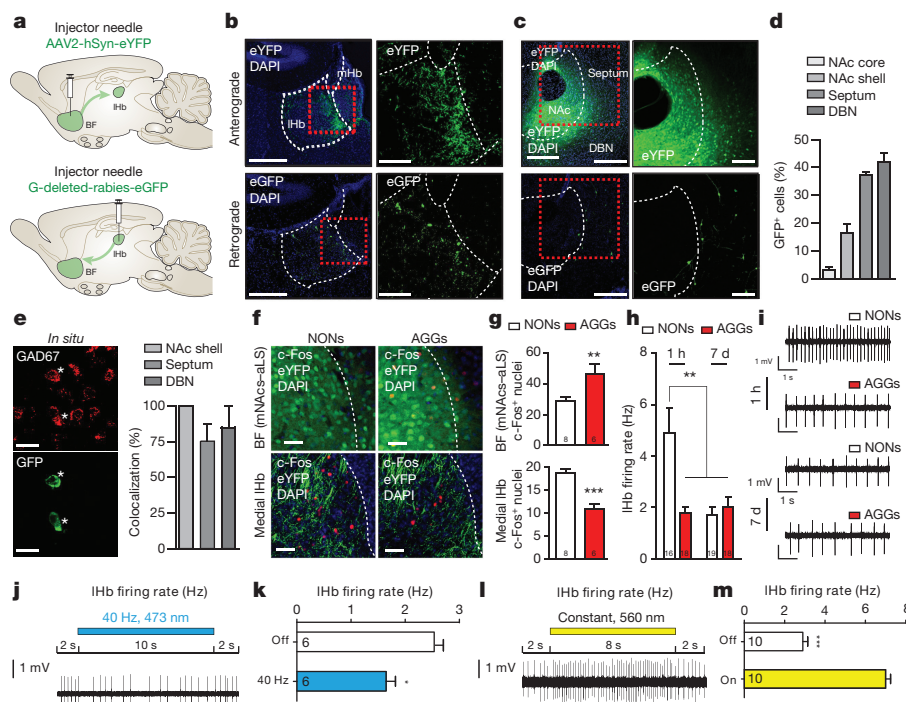


Figure 2 | GABAergic BF-IHb circuit is differentially activated by intruder interactions. **a**, Schematic of anterograde and retrograde tracing strategies. **b**, Representative anterograde AAV2-hSyn-eYFP infections (top, terminals) or retrograde G-deleted-rabies-eGFP infections (bottom, infection site) in IHb. Scale bars: 500 μ m; insets, 150 μ m. **c**, Representative anterograde AAV2-hSyn-eYFP infections (top, infection site) or retrograde G-deleted-rabies-eGFP infections (bottom, cell bodies) in the BF. Scale bars: 400 μ m; insets, 200 μ m. **d**, Percentage retrograde-labelled eGFP⁺ neurons within subnuclei of the anterior BF ($n = 3$ mice, ~229 cells per mouse). **e**, Representative *in situ* hybridization colocalized GAD67 and eGFP in DBN (left) and quantification (right) within the BF ($n = 3$ mice, 14 cells per mouse). Scale bars: 20 μ m. **f**, Representative images of AAV2-hSyn-eYFP infection and c-Fos immunoreactivity in medial NAc shell transition zone of the BF (top) and medial IHb terminals (bottom). Scale bars: 30 μ m. **g**, Quantification of c-Fos immunoreactivity in the medial NAc shell-septum transition zone ($t_8 = 2.655$, $*P < 0.05$; two-tailed

increase IHb firing induced CPA to the intruder-paired side (Extended Data Fig. 7h-j). Taken together, these results implicate the IHb as a key modulator of aggression motivational state.

To determine if BF-IHb neuronal activity regulates the initiation or intensity of aggressive behaviour, we used ChR2^{BF→IHb} and NpHR3^{BF→IHb} (Fig. 4a) in AGGs and NONs during home-cage resident-intruder testing (Fig. 4b). Neither activation nor inhibition of BF-IHb terminals resulted in the initiation of aggressive behaviour (Fig. 4c, d), nor did it modulate social (Fig. 4e) and non-social (Fig. 4f) exploratory behaviours in NON mice. Similarly, AGG::ChR2^{BF→IHb} stimulation failed to initiate immediate attack behaviour, as indexed by no change in attack latency (Fig. 4g). However, AGG::ChR2^{BF→IHb} and AGG::NpHR3^{BF→IHb} stimulation bi-directionally modulated the severity of the aggressive behaviour relative to each other, although a nonsignificant trend was observed when either were compared to AGG::GFP^{BF→IHb} (Fig. 4h). As observed in NONs, AGG::ChR2^{BF→IHb} and AGG::NpHR3^{BF→IHb} photostimulation failed to alter either social (Fig. 4i) or non-social (Fig. 4j) exploratory behaviours. These data indicate that the BF-IHb circuit is important in modulating the intensity of aggressive behaviour; however, it is not a traditional attack initiation area.

On the basis of these data, we hypothesized that the BF-IHb circuit acts in other affective behavioural domains. We performed a behavioural battery to measure non-social generalized anxiety states and reward in naive CD-1 mice (Extended Data Fig. 8a). Both ChR2^{BF→IHb} and

unpaired t -test, $n = 6-8$ mice per group, 3 slices per mouse) and medial IHb ($t_{12} = 5.678$, $***P < 0.001$; two-tailed unpaired t -test, $n = 6-8$ mice per group, 3 slices per mouse). **h**, Firing rate of IHb neurons in AGG and NON mice at 1 h or 7 days after intruder interaction ($F_{1,67} = 10.56$, two-way ANOVA $P < 0.05$; post-hoc test, $**P < 0.01$; $n = 16-19$ cells per group, 4-5 mice per group). **i**, Representative trace of IHb *in vitro* cell-attached firing rates. **j**, Representative trace of IHb *in vitro* cell-attached firing rates during ChR2^{BF→IHb} photostimulation. **k**, Average firing rates of IHb neurons during ChR2^{BF→IHb} ($t_{10} = 3.679$, $**P < 0.01$; two-tailed unpaired t -test, $n = 6$ cells). **l**, Representative trace of IHb *in vitro* cell-attached firing rates during NpHR3^{BF→IHb} photostimulation. **m**, Average firing rates of IHb neurons during NpHR3^{BF→IHb} ($t_{18} = 11.68$, $***P < 0.0001$; two-tailed unpaired t -test, $n = 10$ cells) photostimulation. Data are represented as mean \pm s.e.m. aLS, anterior lateral septum; DAPI, 4',6-diamidino-2-phenylindole; mNAcs, medial nucleus accumbens shell. Experiments were conducted once; n indicates biological replicates.

NpHR3^{BF→IHb} terminal photostimulation failed to modulate anxiety-like behaviours in the open field (Extended Data Fig. 8b, c) and elevated plus maze tasks (Extended Data Fig. 8d, e). However, ChR2^{BF→IHb} stimulation potentiates the rewarding effects of cocaine by increasing the amount of time spent in the cocaine-paired chamber (Extended Data Fig. 8f). Therefore, while the BF-IHb circuit does not influence a generalized anxiety phenotype in the absence of social context or other stimuli, it does generalize to non-social rewarding stimuli such as cocaine.

Our results show individual differences in the rewarding properties of aggressive social interaction that are mediated by the BF-IHb circuit. When exposed to an intruder, AGGs exhibit increased activity of the BF and a concomitant reduction in IHb neuronal firing relative to NONs, contributing to a behavioural preference for environmental contexts associated with the interaction. Importantly, this circuit is not sufficient to induce attack behaviour. Although anatomical studies have identified this BF projection to the IHb in mice and rats¹⁶⁻¹⁸, and diffusion tensor imaging (DTI) suggests probabilistic tract connections between the BF and IHb in humans¹⁹, this is the first study to provide functional evidence that GABAergic BF projections produce inhibitory control of IHb neurons to regulate the valence of aggressive intruder-based interactions. Stimulation or inhibition of BF-IHb projections is both sufficient and necessary to alter the positive or negative valence of an intruder-paired context. Our findings advance the understanding of IHb function in a behaviourally relevant animal model of aggression

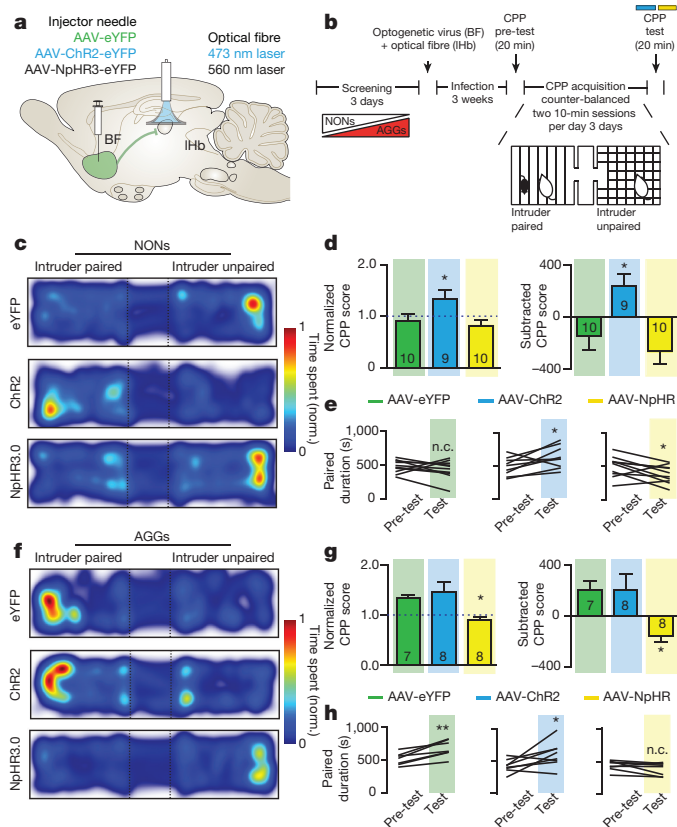


Figure 3 | BF-IHb activity bi-directionally modulates aggression reward. **a, b**, Schematic of optogenetic viral infection strategy (**a**) and aggression CPP procedure (**b**). **c, f**, Representative CPP heatmaps for eYFP^{BF→IHb}, ChR2^{BF→IHb} and NpHR3^{BF→IHb} between NON (**c**) and AGG (**f**) mice. norm., normalized. **d**, Normalized ($F_{2,26} = 5.019$, one-way ANOVA $P < 0.05$; post-hoc test, $*P < 0.05$, $n = 9-10$ per group) and subtracted CPP score ($F_{2,26} = 6.666$, one-way ANOVA $P < 0.01$; post-hoc test, $*P < 0.05$, $n = 9-10$ per group) in NON::eYFP^{BF→IHb}, NON::ChR2^{BF→IHb} and NON::NpHR3^{BF→IHb} mice. **e**, Individual duration in intruder-paired context for NON::eYFP^{BF→IHb} mice ($t_9 = 0.9129$, $P > 0.05$; two-tailed paired t -test, $n = 10$ per group), NON::ChR2^{BF→IHb} mice ($t_8 = 2.362$, $*P < 0.05$; two-tailed paired t -test, $n = 9$ per group), and NON::NpHR3^{BF→IHb} mice ($t_9 = 2.344$, $*P < 0.05$; two-tailed paired t -test, $n = 10$ per group) during the pre-test and test sessions. **g**, Normalized ($F_{2,20} = 5.470$, one-way ANOVA $P < 0.05$; post-hoc test, $*P < 0.05$, $n = 7-8$ per group) and subtracted CPP score ($F_{2,20} = 4.964$, one-way ANOVA $P < 0.05$; post-hoc test, $*P < 0.05$, $n = 7-8$ per group) for intruder-paired context in AGG::eYFP^{BF→IHb}, AGG::ChR2^{BF→IHb} and AGG::NpHR3^{BF→IHb} mice. **h**, Individual duration in intruder-paired context for AGG::eYFP^{BF→IHb} mice ($t_6 = 5.070$, $**P < 0.01$; two-tailed paired t -test, $n = 7$ per group), AGG::ChR2^{BF→IHb} mice ($t_7 = 2.394$, $*P < 0.05$; two-tailed paired t -test, $n = 8$ per group), and AGG::NpHR3^{BF→IHb} mice ($t_7 = 1.763$, $P > 0.05$; two-tailed paired t -test, $n = 8$ per group). Summary data are represented as mean \pm s.e.m. n.c., no change. Experiments were conducted once; n indicates biological replicates.

motivation and provide further understanding into the physiology and neural circuitry of aggression and reward-related behaviours.

While numerous functions have been ascribed to IHb²⁰ neuronal activity, including anxiety²¹, addiction²² and depression²³, there is a noticeable paucity of functional data addressing the role of IHb inputs, outside of those originating from the VTA region, within any of these behavioural domains. Indeed, anatomical tracing experiments have highlighted the complexity of IHb afferents²⁴ and efferents²⁵. With regard to the BF, the lateral septum, DBN and medial NAc shell, but not core, are known to send projections to the IHb^{24,26}. Our study implicates the septo-accumbal transition zone of BF as a critical source of GABAergic tone to the IHb within the context of motivated behaviour. However, on the basis of the fact that these BF GABAergic inputs to IHb

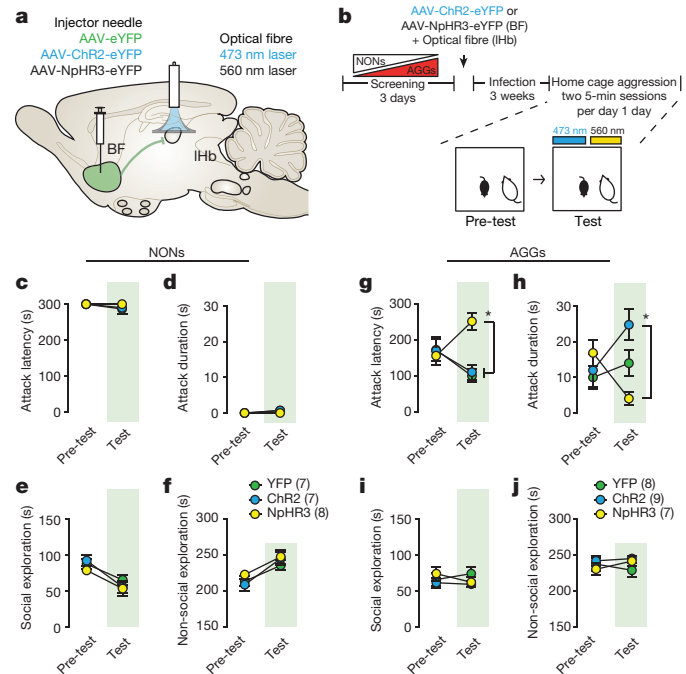


Figure 4 | BF-IHb does not initiate attack but modulates aggression severity. **a, b**, Schematic of optogenetic viral infection strategy (**a**) and aggression procedure (**b**). **c-f**, NON attack latency (**c**), attack duration (**d**), social exploration (**e**) and non-social exploration behaviours (**f**) in pre-test and test sessions (non-significant, $n = 7-8$ per group). **g-j**, AGG attack latency ($F_{2,42} = 6.01$, two-way ANOVA $P < 0.001$, $*P < 0.05$, $n = 7-9$ per group) (**g**), attack duration ($F_{2,42} = 5.666$, two-way ANOVA $P < 0.001$, $*P < 0.05$, $n = 7-9$ per group) (**h**), social exploration (**i**) and non-social exploration behaviours in pre-test and test sessions (**j**). Experiments were conducted once; n indicates biological replicates.

exhibit high tonic activation in acute slice that can be rapidly inhibited by terminal inhibition with NpHR3 (Fig. 2l, m), it is unlikely that they are NAc medium spiny neurons. Finally, based on both *in vitro* and *in vivo* electrophysiological studies, as well as anatomical tracing, we note that there may be a small subset of cells in the BF that either release an excitatory neurotransmitter or act indirectly on the IHb via di-synaptic inputs. It will be interesting in the future to determine what role these neurons have in reward processing.

Our results may provide important information to clinical studies identifying novel targets of deep brain stimulation in the treatment of neuropsychiatric conditions that present with aggression co-morbidity such as substance abuse²⁷ and depression²⁸. Deep brain stimulation protocols within specific BF nuclei²⁹ and the IHb³⁰ have been successfully used to treat intractable major depressive disorder, which is associated with symptoms of increased aggression in men²⁸. Overall, our findings demonstrate a previously unidentified functional role for the IHb and its inputs from the BF in mediating the rewarding component of aggression, and suggest that targeting shared underlying deficits in motivational circuitry may provide useful information for the development of novel therapeutic strategies for treating aggression-related neuropsychiatric disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April 2015; accepted 17 May 2016.

- Anderson, D. J. Optogenetics, sex, and violence in the brain: implications for psychiatry. *Biol. Psychiatry* **71**, 1081–1089 (2012).
- Decety, J., Michalska, K. J., Akitsuki, Y. & Lahey, B. B. Atypical empathic responses in adolescents with aggressive conduct disorder: a functional MRI investigation. *Biol. Psychol.* **80**, 203–211 (2009).
- Yang, C. F. et al. Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell* **153**, 896–909 (2013).

4. Wasman, M. & Flynn, J. P. Directed attack elicited from hypothalamus. *Arch. Neurol.* **6**, 220–227 (1962).
5. Lin, D. *et al.* Functional identification of an aggression locus in the mouse hypothalamus. *Nature* **470**, 221–226 (2011).
6. Unger, E. K. *et al.* Medial amygdalar aromatase neurons regulate aggression in both sexes. *Cell Reports* **10**, 453–462 (2015).
7. Yu, Q. *et al.* Optogenetic stimulation of DAergic VTA neurons increases aggression. *Mol. Psychiatry* **19**, 635 (2014).
8. Takahashi, A. & Miczek, K. A. Neurogenetics of aggressive behavior: studies in rodents. *Curr. Top. Behav. Neurosci.* **17**, 3–44 (2013).
9. Kudryavtseva, N. N., Bakshtanovskaya, I. V. & Koryakina, L. A. Social model of depression in mice of C57BL/6J strain. *Pharmacol. Biochem. Behav.* **38**, 315–320 (1991).
10. Golden, S. A., Covington, H. E. III, Berton, O. & Russo, S. J. A standardized protocol for repeated social defeat stress in mice. *Nat. Protocols* **6**, 1183–1191 (2011).
11. Miczek, K. A., DeBold, J. F. & Thompson, M. L. Pharmacological, hormonal, and behavioral manipulations in analysis of aggressive behavior. *Prog. Clin. Biol. Res.* **167**, 1–26 (1984).
12. Glenn, A. L. & Yang, Y. The potential role of the striatum in antisocial behavior and psychopathy. *Biol. Psychiatry* **72**, 817–822 (2012).
13. Zahm, D. S., Parsley, K. P., Schwartz, Z. M. & Cheng, A. Y. On lateral septum-like characteristics of outputs from the accumbal hedonic “hotspot” of Pecifia and Berridge with commentary on the transitional nature of basal forebrain “boundaries”. *J. Comp. Neurol.* **521**, 50–68 (2013).
14. Callaway, E. M. & Luo, L. Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *J. Neurosci.* **35**, 8979–8985 (2015).
15. Lammel, S. *et al.* Input-specific control of reward and aversion in the ventral tegmental area. *Nature* **491**, 212–217 (2012).
16. Herkenham, M. & Nauta, W. J. Afferent connections of the habenular nuclei in the rat. A horseradish peroxidase study, with a note on the fiber-of-passage problem. *J. Comp. Neurol.* **173**, 123–145 (1977).
17. Sutherland, R. J. The dorsal diencephalic conduction system: a review of the anatomy and functions of the habenular complex. *Neurosci. Biobehav. Rev.* **6**, 1–13 (1982).
18. Lecca, S., Meye, F. J. & Mameli, M. The lateral habenula in addiction and depression: an anatomical, synaptic and behavioral overview. *Eur. J. Neurosci.* **39**, 1170–1178 (2014).
19. Shelton, L. *et al.* Mapping pain activation and connectivity of the human habenula. *J. Neurophysiol.* **107**, 2633–2648 (2012).
20. Hikosaka, O. The habenula: from stress evasion to value-based decision-making. *Nature Rev. Neurosci.* **11**, 503–513 (2010).
21. Lee, E. H. & Huang, S. L. Role of lateral habenula in the regulation of exploratory behavior and its relationship to stress in rats. *Behav. Brain Res.* **30**, 265–271 (1988).
22. Maroteaux, M. & Mameli, M. Cocaine evokes projection-specific synaptic plasticity of lateral habenula neurons. *J. Neurosci.* **32**, 12641–12646 (2012).
23. Li, B. *et al.* Synaptic potentiation onto habenula neurons in the learned helplessness model of depression. *Nature* **470**, 535–539 (2011).
24. Yetnikoff, L., Cheng, A. Y., Lavezzi, H. N., Parsley, K. P. & Zahm, D. S. Sources of input to the rostromedial tegmental nucleus, ventral tegmental area, and lateral habenula compared: a study in rat. *J. Comp. Neurol.* **523**, 2426–2456 (2015).
25. Quina, L. A. *et al.* Efferent pathways of the mouse lateral habenula. *J. Comp. Neurol.* **523**, 32–60 (2015).
26. Felton, T. M., Linton, L., Rosenblatt, J. S. & Morell, J. I. First and second order maternal behavior related afferents of the lateral habenula. *Neuroreport* **10**, 883–887 (1999).
27. Beck, A., Heinz, A. J. & Heinz, A. Translational clinical neuroscience perspectives on the cognitive and neurobiological mechanisms underlying alcohol-related aggression. *Curr. Top. Behav. Neurosci.* **17**, 443–474 (2014).
28. Martin, L. A., Neighbors, H. W. & Griffith, D. M. The experience of symptoms of depression in men vs women: analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry* **70**, 1100–1106 (2013).
29. Bewernick, B. H., Kayser, S., Sturm, V. & Schlaepfer, T. E. Long-term effects of nucleus accumbens deep brain stimulation in treatment-resistant depression: evidence for sustained efficacy. *Neuropsychopharmacology* **37**, 1975–1985 (2012).
30. Sartorius, A. & Henn, F. A. Deep brain stimulation of the lateral habenula in treatment resistant major depression. *Med. Hypotheses* **69**, 1305–1308 (2007).

Acknowledgements This research was supported by US National Institutes of Health grants R01 MH090264, P50 MH096890 and P50 AT008661-01 (S.J.R.), R01 MH092306 (M.H.H.), T32 MH087004 (M.L.P., M.H. and M.F.), T32 MH096678 (M.L.P.), F30 MH100835 (M.H.), F31 MH105217 (M.L.P.), National Institute of General Medical Sciences 1F12GM117583-01 (S.A.G.) and the National Natural Science Foundation of China 81200862 (H.Z.). We would like to thank K. Miczek and Y. Shaham for their input.

Author Contributions S.A.G. and S.J.R. designed and wrote the manuscript. S.A.G., D.J.C., M.H., C.M., J.J.W., M.L.P., N.R. H.A., G.E.H., M.F., D.B., L.K., J.T. and B.K. collected behavioural and immunohistochemistry data and aided in data analysis. H.Z., M.-H.H., D.C., K.G. and M.L.S. designed, carried out and analysed electrophysiological experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.R. (scott.russo@mssm.edu).

Reviewer Information *Nature* thanks O. Hikosaka and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals. Male CD-1 (ICR) mice (35–45 g, sexually experienced retired breeders; Charles River Laboratories (CRL)) were obtained at 4 months of age. All breeders were confirmed by CRL to have had equal access, experience and success as breeders. Male C57BL/6J mice (20–30 g; The Jackson Laboratory) were obtained at 7–8 weeks of age and used as novel intruders. All mice were allowed 1 week of acclimation to the housing facilities before the start of experiments. CD-1 mice were single housed, and C57BL/6J mice were group housed. All mice were maintained on a 12 h light:dark cycle with *ad libitum* access to food and water. Procedures were performed in accordance with the National Institutes of Health Guide for Care and Use of Laboratory Animals and the Icahn School of Medicine at Mount Sinai Institutional Animal Care and Use Committee.

Aggression screening and ethological analysis. Aggression screening was performed as previously described¹⁰. After a minimum of 1 week habituation to home cages, experimental CD-1 mice were exposed to a novel C57BL/6J intruder for 3 min daily over 3 consecutive days. Each intruder presentation was performed in the home cage of the CD-1 mouse between 1–3 PM daily under white light conditions. During screening sessions the cage top along with feeding and water apparatus were replaced with a clear Plexiglass cover to allow unimpeded viewing and video recording of screening sessions. The duration and number of screening sessions were selected to prevent induction of stress- and anxiety-related behaviours in CD-1 mice (Extended Data Tables 1 and 2), which has been shown to occur during extended antagonist encounters³¹. This allows for separation between aggression and stress-related states. All screening sessions were video recorded for later ethological analysis using a digital colour video camera. Ethological analysis of aggression behaviour was performed by two blinded observers, recording (1) latency to initial aggression, (2) the number of aggressive bouts, (3) the total duration of aggression, and (4) the mean duration of aggressive bouts. Operational definitions for these behaviours are defined as follows: initiation of aggression is defined by the first clear physical antagonist interaction initiated by the CD-1 mouse, not including grooming or pursuit behaviour. Aggressive bouts are defined by cycles of initiated aggression with continuous orientation by the CD-1 mouse towards the intruder, and only defined as completed when the CD-1 mouse has physically reoriented away from the intruder. This definition allows for slight breaks (less than 5 s) in continuous physical interaction within an aggressive bout, assuming the CD-1 mouse has remained oriented towards the intruder throughout. CD-1 mice were defined as AGG if they initiated aggression during any of the three screening sessions and NON were defined as those that showed no aggression during any screening sessions. All aggression screening was halted if an intruder showed any signs of injury in accordance with our previously published protocol¹⁰.

Aggression CPP and behavioural analysis. The aggression CPP protocol, developed on the basis of a previously published cocaine CPP protocol³², consisted of three phases: pre-test, acquisition, and test. Mice were acclimated to the testing facility for 1 h before testing. All phases were conducted under red light and sound-attenuated conditions. The CPP apparatus (Med Associates) consisted of two unique conditioning chambers with a neutral middle zone that allowed for unbiased entry into either conditioning chamber at the initiation of each trial. All CPP sessions were video recorded using Noldus Ethovision 3.0 (Noldus Information Technology). During the pre-test phase, mice were placed into the middle chamber of the conditioning apparatus and allowed to freely explore the full extent of the CPP apparatus for 20 min. There were no group differences in bias for either chamber, and conditioning groups were balanced in an unbiased fashion to account for start side preference. The acquisition phase consisted of three successive days with two conditioning trials each day for a total of six acquisition trials. Morning trials (between 8:00 AM and 10:00 AM) and afternoon trials (between 3:00 PM and 5:00 PM) consisted of CD-1 mice confined to one chamber for 10 min while in the presence or absence of a novel C57BL/6J intruder. All groups were counterbalanced for conditioning chamber. A total of three conditioning trials to the intruder-paired and intruder-unpaired context were performed. On the test day, CD-1 mice were placed into the middle arena of the CPP apparatus without an intruder and allowed to freely explore both chambers for 20 min. Analysis of duration spent within either context was used to identify a CPP or CPA to the intruder-paired context. For optogenetic experiments, stimulation was performed during the full duration of the test phase. Total locomotor responses were also assessed to ensure equal exploratory behaviour between groups. Behavioural analysis of aggression CPP data was performed by assessing (1) normalized CPP (test phase duration spent in the intruder-paired chamber divided by the pre-test duration spent in the intruder-paired chamber, accounting for behaviour during both sessions), (2) subtracted CPP (test phase duration spent in the intruder-paired chamber minus test phase duration spent in the intruder-unpaired chamber, accounting for test session behaviour only), and (3) group and individual durations in both pre-test and test sessions.

Sensory CPP. Sensory CPP was performed and analysed identically to the aggression CPP procedure, with the exception that the intruder C57BL/6J was placed within a physical barrier to provide only sensory contact with the resident CD-1 mice.

Cocaine CPP. A previously published cocaine CPP protocol³² was used, which consisted of three phases: pre-test, acquisition, and test. Mice were acclimated to the testing facility for 1 h before testing. All phases were conducted under red light and sound-attenuated conditions. The CPP apparatus (Med Associates) consisted of two unique conditioning chambers with a neutral middle zone that allowed for unbiased entry into either conditioning chamber at the initiation of each trial. All CPP sessions were video recorded using Noldus Ethovision 3.0 (Noldus Information Technology). During the pre-test phase, mice were placed into the middle chamber of the conditioning apparatus and allowed to freely explore the full extent of the CPP apparatus for 30 min. There were no group differences in bias for either chamber, and conditioning groups were balanced in an unbiased fashion to account for start side preference. The acquisition phase consisted of two successive days with two conditioning trials each day for a total of four acquisition trials. Morning trials (between 8:00 AM and 10:00 AM) and afternoon trials (between 3:00 PM and 5:00 PM) consisted of CD-1 mice confined to one chamber for 20 min paired with an intraperitoneal injection of cocaine (10 mg kg⁻¹); afternoon sessions were paired with saline injections. All groups were counterbalanced for conditioning chamber. On the test day, CD-1 mice were placed into the middle arena of the CPP apparatus and allowed to freely explore both chambers for 30 min. Analysis of duration spent within either context was used to identify a CPP or CPA to the cocaine-paired context. For optogenetic experiments, stimulation was performed during the full duration of the test phase. Total locomotor responses were also assessed to ensure equal exploratory behaviour between groups.

Sucrose preference. Sucrose preference was performed as previously described³³. One week after the final screening session, AGG and NON mice had their standard water bottle removed and replaced with two 50-ml conical tubes with sipper tops filled with water. After a 24-h habituation period, water from one 50-ml conical tube was replaced with 1% sucrose. All tubes were weighed, and mice were allowed 24 h to drink. Tubes were then reweighed, and their locations in the wire tops were switched before a second 24-h period of drinking. At the end of the second day of sucrose testing, preference was calculated as the total amount of sucrose consumption divided by the total amount of fluid consumed over the 2 days of sucrose availability.

Elevated plus maze. The elevated plus maze was performed as previously described³⁴. One week after the final screening session, AGG and NON mice were acclimated to the testing facility for 1 h before testing and then placed in the elevated plus maze under red light conditions for 5 min. Each arm of the maze measured 12 × 50 cm. The Plexiglas cross-shaped maze consisted of two open arms with no walls and two closed arms (40-cm-high walls) and was on a pedestal 1 m above floor level. Behaviour was tracked using an automated system (Noldus Ethovision; Noldus Interactive technologies). Behaviour was measured as total time in open and closed arms.

Open field and locomotor measures. Open field was performed as previously described³⁴. One week after the final screening session, AGG and NON mice were acclimated to the testing facility for 1 h before testing. Open-field testing was performed in black Plexiglas arenas (42 × 42 × 42 cm; Nationwide Plastics) under red light conditions. Testing sessions were 10 min long. Behaviour was tracked using an automated system (Noldus Ethovision; Noldus Interactive technologies) to record the total distance moved and time spent in the total arena and a delineated 'centre zone' (24 cm × 24 cm).

Forced-swim test. The forced-swim test was performed as previously described³⁴. One week after the final screening session, AGG and NON mice were placed in the test room for an hour before behavioural testing for habituation. Mice were tested in a 4-litre Pyrex glass beaker, containing 2 litres of water at 25 ± 1 °C for 6 min. Behaviour was videotaped (Noldus Ethovision; Noldus Interactive technologies) and analysed for duration immobile, duration mobile and total movement.

Social interaction (approach). Social approach testing was performed as previously described¹⁰. One week after the final screening session, AGG and NON mice were acclimated to the testing facility for 1 h before testing, and all testing was performed under red light conditions. Mice were placed in an open field black Plexiglas arena (42 × 42 × 42 cm; Nationwide Plastics) with a small animal cage placed at one end. Their movements were then automatically monitored and recorded (Ethovision 3.0; Noldus Information Technology) for 2.5 min in the absence (target absent phase) of a social target. This phase is used to determine baseline exploratory behaviour. We then immediately measured 2.5 min of exploratory behaviour in the presence of a caged novel CD-1 or C57BL/6J mouse (target present phase), again recording total distance travelled and duration of time spent in the interaction and corner zones. Social interaction behaviour is determined by the total time spent in each zone.

Novel object versus social target preference. The novel object versus social target test consisted of two phases: pre-test and test on consecutive days, as previously described³⁵. One week after the final screening session, AGG and NON mice were acclimated to the testing facility for 1 h before both sessions. All phases were run under red-light and sound-attenuated conditions. The testing apparatus (Med Associates) consisted of two identical chambers, with a neutral middle zone that allowed for unbiased entry into either chamber at the initiation of each trial. All sessions were video recorded from above (Noldus Ethovision 3.0, Noldus Information Technology) for later behavioural analysis. Briefly, during the pre-test phase, mice were placed into the middle chamber of the apparatus and allowed to freely explore all zones for 5 min. There were no group differences in pre-test preference for either chamber. Conditioning groups were then balanced in an unbiased way to account for individual animals' preference. On the test day, mice were placed back into the apparatus in the presence of both a novel object (an upside-down steel-bar pencil holder) on one side and a social target (identical pencil holder containing either a novel CD-1 or a C57BL/6J mouse) on the other. Test mice were allowed to freely explore the apparatus for 5 min. The time spent in each chamber was recorded and used for analysis. The subtracted social score is derived by subtracting time in social-paired chamber from time in novel object-paired chamber during the test phase. Normalized social score is the ratio of time spent in the chamber of interest (social target or novel object) during the test phase over the pre-test phase.

Blood sampling and testosterone/corticosterone ELISA. Submandibular vein bleeds were taken from mice 4–24 h after the final screening session as previously described³⁶. Serum testosterone (RND Systems, Testosterone Parameter Assay Kit) and corticosterone (Immunodiagnostic Systems, IDS Corticosterone EIA Kit) levels were assessed via ELISA according to manufacture specifications. Briefly, blood was collected in a serum separator tube, allowed to clot for 30 min at room temperature, and centrifuged at 1,000g for 15 min and stored frozen (−20°C) until analysis. The sensitivity of the testosterone assay (minimum detectable dose ranging from 0.012 to 0.041 ng ml^{−1}) falls well below the ranges detected experimentally within our cohort (lowest serum testosterone concentration of 2.58 ng ml^{−1}).

Perfusion and brain tissue processing. For immunohistochemistry and histology, mice were given a euthanizing dose of 15% chloral hydrate and transcardially perfused with cold 1% paraformaldehyde in PBS (pH 7.4) followed by fixation with cold 4% paraformaldehyde in PBS. Brains were dissected and post-fixed for 18 h in the same fixative. Coronal sections were prepared on a vibratome (Leica) at 50 µm to assess viral placement and immunohistochemistry. For *in situ* hybridization, mice brains were rapidly removed and flash frozen in −30°C isopentane for 60 s and then kept at −80°C until sectioning.

Immunohistochemistry, *in situ* hybridization and confocal microscopy. For c-Fos experiments, sections were incubated overnight in blocking solution (3% normal donkey serum, 0.3% Triton X-100 in PBS), washed in PBS for 2 h, then incubated for 48 h in primary antibody (rabbit anti-c-Fos (Santa Cruz Biotechnology, SC-42) 1:2,000). Slices were then washed in PBS for 2 h, incubated in secondary antibody for 2 h (donkey anti-rabbit Cy2 1:200 (Jackson ImmunoResearch)), then washed in PBS for 30 min before staining with 1 µg ml^{−1} DAPI (Sigma) for 20 min. Sections were then mounted, air-dried overnight and coverslipped with hardset Vectashield (Jackson ImmunoResearch). All slices were images using a Zeiss LSM 780. For c-Fos analysis, all images were taken at ×20 magnification for both the BF and IHB, using the tile-scan function to span the entire region of interest. Analysis of c-Fos-positive nuclei was performed using NIH ImageJ in conjunction with the 'analyze particle' function on single images. For representative images demonstrating the areas of viral infection, images were acquired at ×10 magnification using the tile-scan function.

For all other immunohistochemistry, coronal sections (50 µm) were used for all immunofluorescence experiments. Sections were incubated in blocking solution (3% normal donkey serum, 0.3% Triton X-100 in PBS) for 1 h. Sections were then incubated in primary antibody overnight at 4°C (VGAT 1:500 (Synaptic Systems); VGLUT1 (Millipore) 1:5,000; 1:250; GFP (Aves) 1:1,000). Next, sections were washed in PBS for 60 min and then incubated in secondary antibody for 2 h (donkey anti-guinea pig Cy5 1:400; donkey anti-goat Cy5 1:400; donkey anti-chicken Cy2 1:400 (Jackson ImmunoResearch)), then washed with PBS for 60 min, stained with 1 µg ml^{−1} DAPI (Sigma) for 20 min, mounted and air-dried overnight. Slides were quickly washed in ethanol 70%, 95%, 100% and Citrosolv (Fisher), and coverslipped with DPX mounting medium (Electron Microscopy Sciences). All slices were images using a Zeiss LSM 780. For puncta imaging, 1-µm z-stacks were taken at ×100 magnification for both the BF and IHB. Deconvolution was performed on all z-stacks with AutoQuant X (MediaCybernetics). For representative image demonstrating the area of viral infection, images were acquired at ×100 magnification using the tile-scan function.

For *in situ* hybridization, RNAscope Multiplex Fluorescent Kits (Advanced Cell Diagnostics) were used with the company-provided procedure. Briefly, fresh frozen brains were slide mounted at 16 µm thickness, fixed for 15 min in cold 4% PFA, seri-

ally dehydrated with increasing EtOH concentration washes (50%, 75% 100% EtOH for 2 min each), and pre-treated with protease reagent (Protease IV, RNAscope) for 20 min. Proprietary probes (Advanced Cell Diagnostics) for eGFP (Channel 1) or GAD67 (Channel 2) were hybridized at 40°C for 2 h, and then subjected to a series of amplification steps at 40°C (1-FL: 30 min; 2-FL: 15 min; 3-FL: 30 min; 4-FL: 15 min). For the fourth amplification step, Reagent Alt-A was used, corresponding with Channel 1 visualization at 488 nm and Channel 2 at 550 nm. Finally slides were treated for 2 min with DAPI, an immediately coverslipped with EcoMount.

***In vitro* electrophysiology.** All recordings were performed blind to experimental condition, and performed in both NON and AGG CD-1 mice. For optogenetic slice electrophysiology, mice were anaesthetized with isoflurane, and perfused with cold artificial cerebrospinal fluid (aCSF) composed of (in mM): 128 NaCl, 3 KCl, 1.25 NaH₂PO₄, 10 D-glucose, 24 NaHCO₃, 2 CaCl₂ and 2 MgCl₂ (oxygenated with 95% O₂ and 5% CO₂, pH 7.35, 295–305 mOsm) as described in our previous work^{37,38}. Briefly, acute brain slices containing the IHB were cut using a microslicer (DTK-1000, Ted Pella) in 95% O₂ and 5% CO₂ saturated sucrose-aCSF, which was derived by fully replacing NaCl with 254 mM sucrose. Slices were maintained in the holding chamber containing aCSF for 1 h at 37°C. Slices were then transferred into a recording chamber fitted with a constant flow rate of aCSF equilibrated with 95%/5% O₂/CO₂ (2.5 ml min^{−1}) maintained at 35°C. Cell-attached recording mode was used to measure the firing rates of IHB neurons. In these recording experiments, glass recording pipettes (7–10 MΩ) were filled with an internal solution composed of (in mM): 115 potassium gluconate, 20 KCl, 1.5 MgCl₂, 10 phosphocreatine, 10 HEPES, 2 magnesium ATP and 0.5 GTP (pH 7.2, 285 mOsm). For the experiments to measure inhibitory postsynaptic currents, whole-cell recordings were performed under voltage-clamp mode (holding at −70 mV) in the presence of kynurenic acid (1 mM) with or without gabazine (2 µM) in aCSF. Glass recording pipettes (3–4 MΩ) for these whole-cell studies were filled with the internal solution composed of (mM): 120 CsCl, 10 phosphocreatine-Na, 10 HEPES, 10 EGTA, 2 ATP-Mg, 0.3 GTP-Tris (pH 7.2, 285 mOsm). Data acquisition was conducted using a Digidata 1440A digitizer and pClamp 10.2 (Axon Instruments).

Stereotaxic surgery and viral gene transfer. All surgeries were performed under aseptic conditions using anaesthetic. Briefly, mice were anesthetized with a mixture of ketamine (100 mg per kg body weight) and xylazine (10 mg per kg body weight) and positioned in a small-animal stereotaxic instrument (David Kopf Instruments) and the skull surface was exposed. Thirty-three-gauge syringe needles (Hamilton Co.) were used to bilaterally infuse either 0.5 µl (BF) or 0.4 µl (IHB) of virus over a 5 min period and the needle was removed after 5 min. NAc shell–septum transition zone BF stereotaxic coordinates taken from bregma (anteroposterior +1.5 mm; mediolateral, +1.6 mm; dorsoventral, −4.4 mm; angle 10°). IHB stereotaxic coordinates taken from bregma (anteroposterior, −1.7 mm; mediolateral, +0.4 mm; dorsoventral, −2.5 mm; angle 0°). For IHB optogenetic experiments, animals were implanted with an optical fibre at the same time as viral injection (dorsoventral, −2.0 mm). For secure fixture of the implantable fibre to the skull, the skull was dried and then industrial-strength dental cement (Grip cement; Dentsply) was added between the base of the implantable fibre and the skull. For non-conditional axonal tract tracing, 0.5 µl AAV2-hSyn-eYFP (1.5 × 10^{11–13} infectious units per ml, UNC Vector Core) was injected bilaterally into the BF. For retrograde tracing, 0.4 µl G-deleted-rabies-eGFP (1.33 × 10⁸ infectious units per ml, Salk Gene Transfer Targeting and Therapeutics Core) was injected into the IHB. For behavioural optogenetic experiments, 0.5 µl of non-conditional AAV2-hSyn-eYFP, AAV2-hSyn-hChR2(H134R)-eYFP or AAV-hSyn-eNpHR3.0-eYFP (1.5 × 10^{11–13} infectious units per ml, UNC Vector Core) were injected into the BF (terminal stimulation) or IHB (cell body stimulation). All non-rabies AAV injections were performed between 4–6 weeks before tracing or behavioural experiments; rabies-infected brains were collected 7 days after injection.

Blue light stimulation. Optical fibres (Thor Labs, BFL37-200) were connected using an FC/PC adaptor to a 473-nm blue laser diode (Crystal Laser, BCL-473-050-M) and a stimulator (Agilent Technologies, no. 33220A) was used to generate blue light pulses. For all *in vivo* behavioural experiments and optrode recordings, mice were given 40 Hz 5 ms light stimulations. Intensity of light delivered to ferrule was ~10 mW. These parameters are consistent with previously validated and published protocols for NAc medium spiny neurons^{39,40} and IHB neurons^{15,41}.

Yellow light stimulation. Optical fibres (Thor Labs, SFS200/220Y) were connected using an FC/PC adaptor to a 561-nm yellow laser diode (Crystal Laser, CL561-050-L), and a stimulator (Agilent Technologies, no. 33220A) was used to generate yellow light pulses. For *in vivo* optrode recordings we tested a protocol of 8 s of yellow light on followed by 2 s of light off. Intensity of light delivered to ferrule was ~10 mW. These parameters are consistent with previously validated and published protocols for NAc MSNs and IHB neurons^{42,43}.

***In vivo* recordings**

Optrode construction and implantation. An optrode was constructed by gluing four tetrodes to an optical fibre. Four tetrodes spun of 12.7-µm-diameter nichrome wire

(California Fine Wire) were glued to a 200- μ m-diameter optical fibre (Thor Labs, SFS200/220Y) and cut so that they extended between 750 and 250 μ m beyond the end of the fibre. The tetrodes were pinned into an electrode interface board (EIB; Neuralynx) and the tips were plated by passing 0.2 μ A current pulses through the individual wires and a gold solution until the impedance reached 150–200 k Ω m. The optrode was mounted on a stereotax arm (Kopf Instruments) and then lowered into the brain during surgery. Two small holes were drilled anterior and posterior to the recording site to serve as sites for ground screws. The ground screws were constructed by soldering stainless steel self-tapping screws to 3 mm stainless steel wire secured to the EIB. Screws were inserted far enough to come in contact with dura.

Recording. Recordings were carried out using a Digital Lynx 16SX recording system and Cheetah data acquisition software (Neuralynx). Signals from the tetrodes were bandpass filtered between 600 and 9,000 Hz and digitized at 32 kHz. Spike detection was performed in real time using a thresholding procedure: when the filtered signal reached threshold amplitude on any wire, a sweep including 8 data points before the crossing and 24 points after (32 points, or 1 ms) were saved as a putative spike event. Spike sorting and noise filtering was performed offline. The laser intensity was adjusted to \sim 5 mW at the tip of the optrode before implantation. The optrode was lowered using the stereotax arm until the tetrode tips reached the dorsal extent of the LHb. Once the tissue and recordings stabilized, the optrode was slowly advanced until spikes were observed on at least one of the tetrodes. Spike amplitude and firing rate were allowed to stabilize and observed for several minutes before recording. For all trials a 30 s baseline recording was acquired, followed by 1 min of stimulation and ending with a 30 s post-stimulation baseline. The optrode was then stepped forward and this procedure repeated until the inferior extent of the LHb was reached.

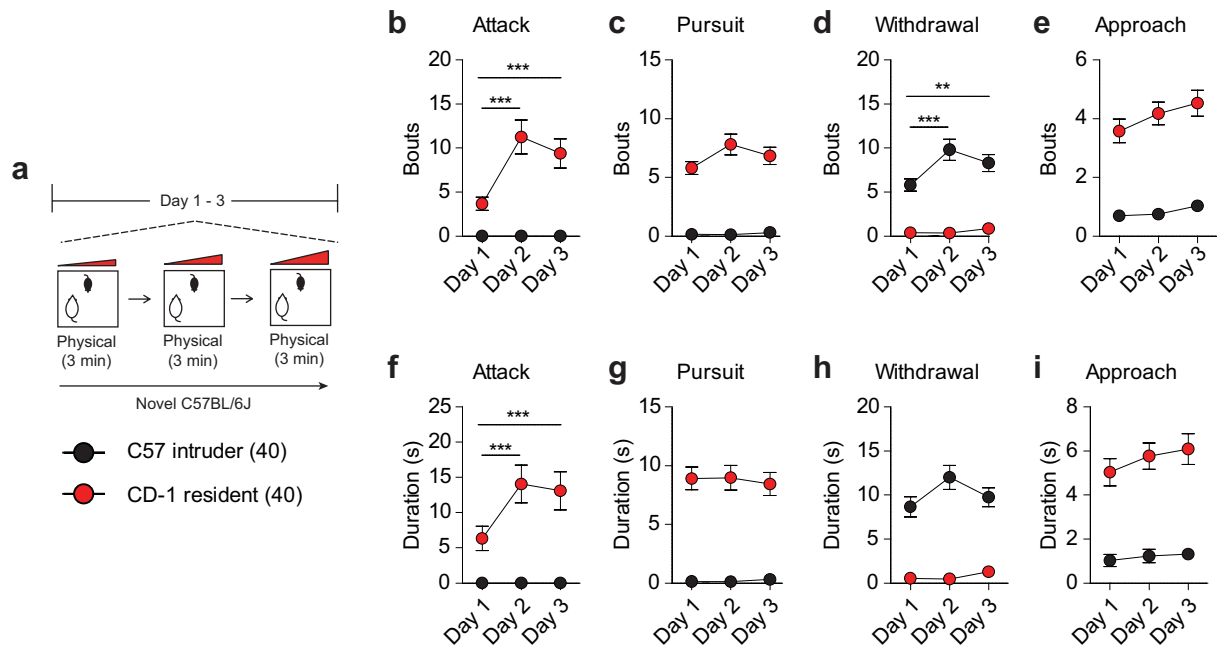
Analysis. Data were analysed using custom scripts written in Matlab (MathWorks). A first round of preliminary spike sorting was carried out using spike waveforms as parameters in KlustaKwik⁴⁴. The output from KlustaKwik was then imported into Matlab and clusters were manually edited using custom spike sorting software. Clearly separated clusters of spikes were assigned to functional units and entered into further analysis; noise spikes (for example, from spurious threshold crossings) and units that fired fewer than 100 spikes during recording were discarded. Spike rates were calculated in 2-s non-overlapping bins across the baseline and stimulation epochs. The resulting functions were smoothed using a Gaussian window with a standard deviation of 10 s. The rate function for each unit was then z-scored across all three epochs. For statistical analyses, rates were calculated in either 15-s bins or bins encompassing the entire baseline and stimulation periods. No smoothing was applied. The rate functions for each unit were z-scored across all three epochs and the z-scored rate functions were used to assess statistical significance.

Randomization and blinding. All experimenters were blinded to experimental condition. Mice were first screened to determine whether they were aggressive or

non-aggressive and then randomly assigned to optogenetic viral conditions for further behavioural analysis. For behavioural studies in Fig. 1 and slice physiology and c-Fos mapping studies in Fig. 2, AGGs and NONs were pre-screened for aggression and assigned to groups on the basis of their behavioural profile.

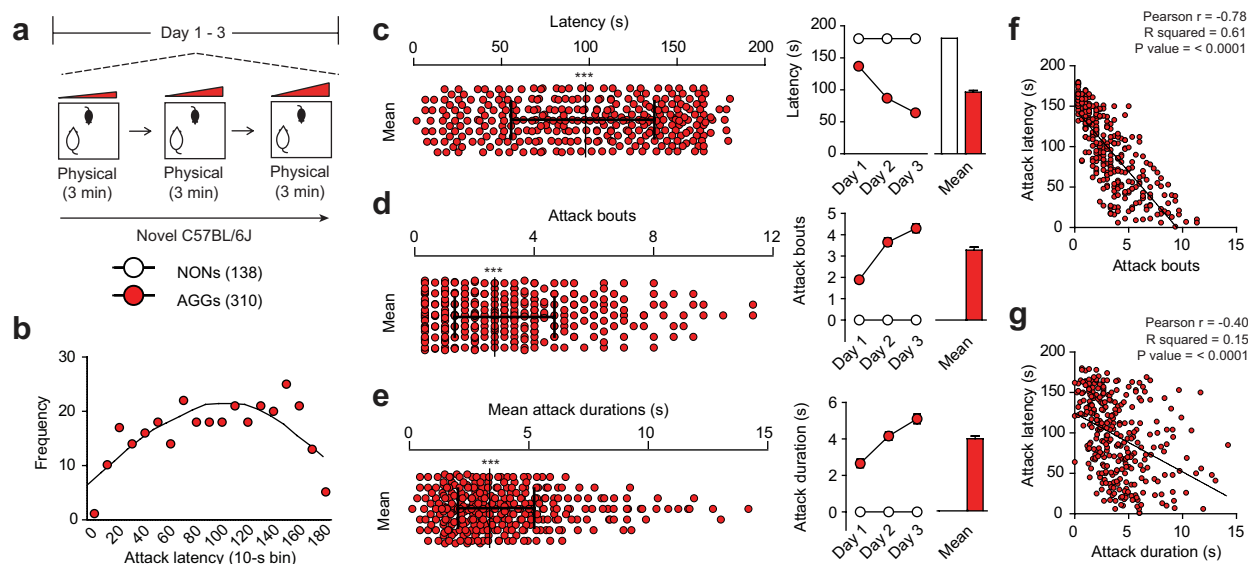
Statistical analysis. Sample size was calculated based on previous studies using Statmate from Graphpad prism (Graphpad Software). All *t*-tests, one-way ANOVAs, two-way ANOVAs and chi-squared tests were performed using Graph Pad Prism software (Graphpad Software Inc.). Bonferroni was used as a post-hoc test when appropriate for one-way and two-way ANOVAs. Normality was determined by D'Agostino–Pearson, Shapiro–Wilk and Kolmogorov–Smirnov normality tests. Statistical significance was set at $P < 0.05$.

31. Kudryavtseva, N. N., Bondar, N. P. & Avgustinovich, D. F. Association between experience of aggression and anxiety in male mice. *Behav. Brain Res.* **133**, 83–93 (2002).
32. Russo, S. J. *et al.* Nuclear factor kappa B signaling regulates neuronal morphology and cocaine reward. *J. Neurosci.* **29**, 3529–3537 (2009).
33. Golden, S. A. *et al.* Epigenetic regulation of RAC1 induces synaptic remodeling in stress disorders and depression. *Nature Med.* **19**, 337–344 (2013).
34. Krishnan, V. *et al.* Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131**, 391–404 (2007).
35. Yang, M., Silverman, J. L. & Crawley, J. N. Automated three-chambered social approach task for mice. *Curr. Protoc. Neurosci.* Chapter 8, Unit 8 26 (2011).
36. Golde, W. T., Gollobin, P. & Rodriguez, L. L. A rapid, simple, and humane method for submandibular bleeding of mice using a lancet. *Lab Anim. (NY)* **34**, 39–43 (2005).
37. Chaudhury, D. *et al.* Rapid regulation of depression-related behaviours by control of midbrain dopamine neurons. *Nature* **493**, 532–536 (2013).
38. Friedman, A. K. *et al.* Enhancing depression mechanisms in midbrain dopamine neurons achieves homeostatic resilience. *Science* **344**, 313–319 (2014).
39. Lobo, M. K. *et al.* Δ FosB induction in striatal medium spiny neuron subtypes in response to chronic pharmacological, emotional, and optogenetic stimuli. *J. Neurosci.* **33**, 18381–18395 (2013).
40. Lobo, M. K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330**, 385–390 (2010).
41. Stamatakis, A. M. & Stuber, G. D. Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. *Nature Neurosci.* **15**, 1105–1107 (2012).
42. Chandra, R. *et al.* Optogenetic inhibition of D1R containing nucleus accumbens neurons alters cocaine-mediated regulation of Tiam1. *Front. Mol. Neurosci.* **6**, 13 (2013).
43. Aquili, L., Liu, A. W., Shindou, M., Shindou, T. & Wickens, J. R. Behavioral flexibility is increased by optogenetic inhibition of neurons in the nucleus accumbens shell during specific time segments. *Learn. Mem.* **21**, 223–231 (2014).
44. Kadir, S. N., Goodman, D. F. M. & Harris, K. D. High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput.* **26**, 2379–2394 (2014).



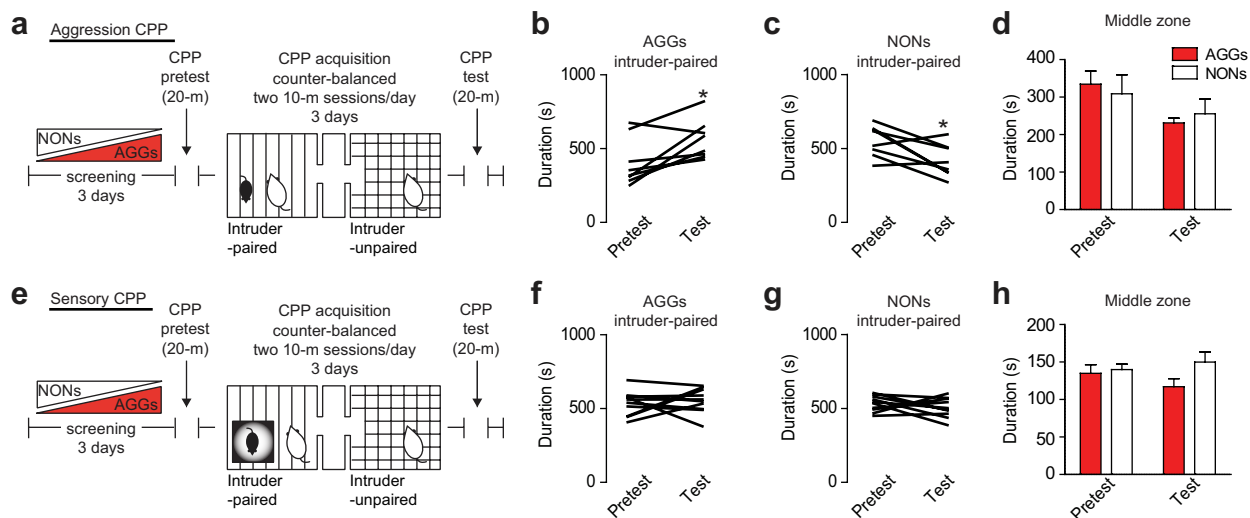
Extended Data Figure 1 | Social behaviours exhibited by resident CD-1 and intruder C57 mice during aggression screening. **a**, Experimental schematic of aggression screening procedure used in a subset (40 residents and 40 intruders) of mice to quantify social behaviours. **b–e**, Bouts of attacks ($F_{2,156} = 13.10$, two-way ANOVA *** $P < 0.0001$; post-hoc test *** $P < 0.001$; $n = 40$ per group) (b), pursuits (c), withdrawals

($F_{2,156} = 5.745$, two-way repeated measures ANOVA ** $P < 0.001$; post-hoc test *** $P < 0.001$; $n = 40$ per group) (d) and non-aggressive social approaches (e). **f–h**, Duration of attacks ($F_{2,156} = 7.069$, two-way repeated measures ANOVA ** $P < 0.001$; post-hoc test *** $P < 0.001$; $n = 40$ per group) (f), pursuits (g), withdrawals (h) and non-aggressive social approaches (e). All data are presented as mean \pm s.e.m.



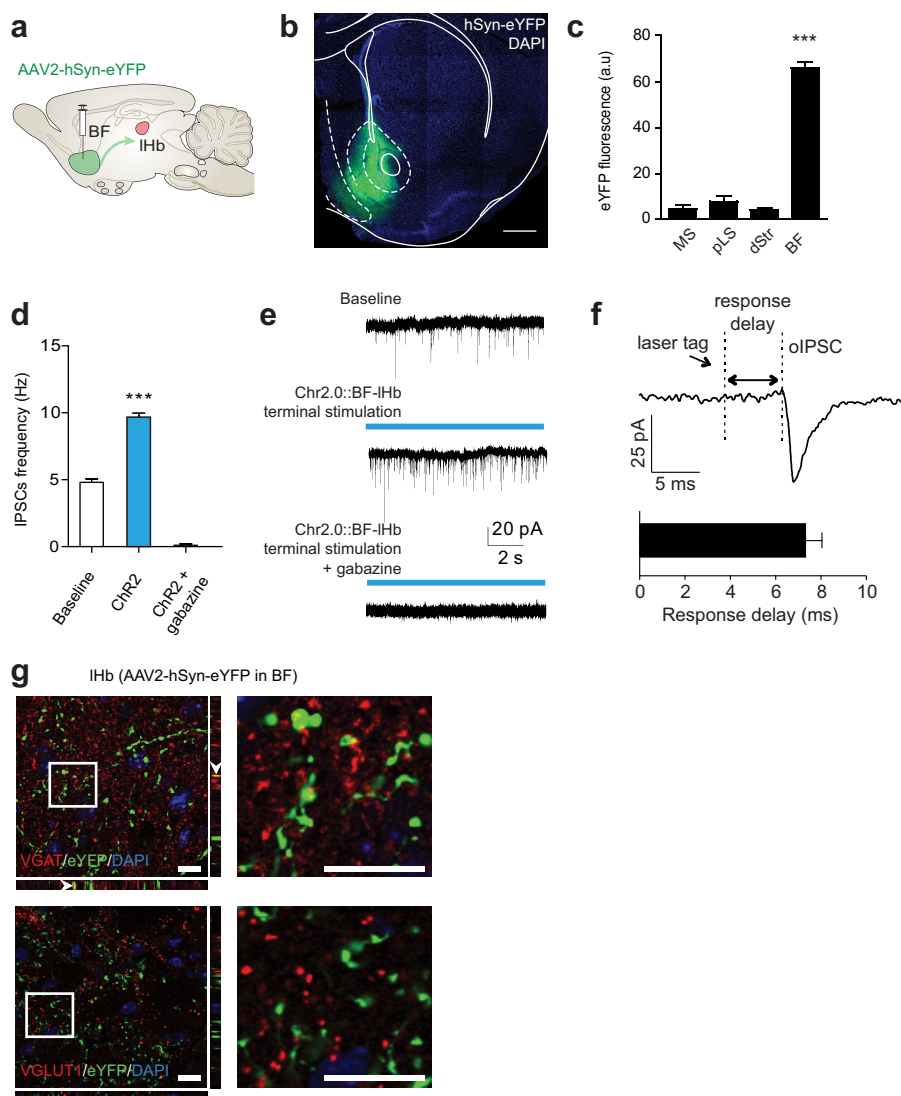
Extended Data Figure 2 | Detailed ethological analysis of AGG aggression-related behaviours. **a**, Experimental schematic of aggression screening procedure used in a sample (448 mice total; 138 NON and 310 AGG) of mice. **b**, Histogram of attack latency frequency using 10-s bins. **c–e**, Mean distribution across screening sessions (left) and individual screening sessions (right) for latency to aggression ($F_{2,1338} = 49.37$, two-way repeated measures ANOVA $P < 0.001$; post-hoc test, $*P < 0.001$; $n = 138–310$) (**c**), number of attack bouts ($F_{2,1338} = 21.03$, two-way repeated measures ANOVA $P < 0.001$; post-hoc test, $*P < 0.001$; $n = 138–310$) (**d**)

and mean attack duration ($F_{2,1338} = 11.96$, two-way repeated measures ANOVA $P < 0.001$; post-hoc test, $*P < 0.001$; $n = 138–310$) (**e**). **f**, **g**, Correlation of mean latency to initial aggression with mean attack bouts ($r = -0.78$, $P < 0.0001$) (**f**) and mean duration of attack bouts ($r = -0.40$, $P < 0.0001$) (**g**). Distribution plots are presented as the median with interquartile range and normality determined by D'Agostino–Pearson, Shapiro–Wilk and Kolmogorov–Smirnov normality tests ($P < 0.0001$). Summary data are represented as mean \pm s.e.m.



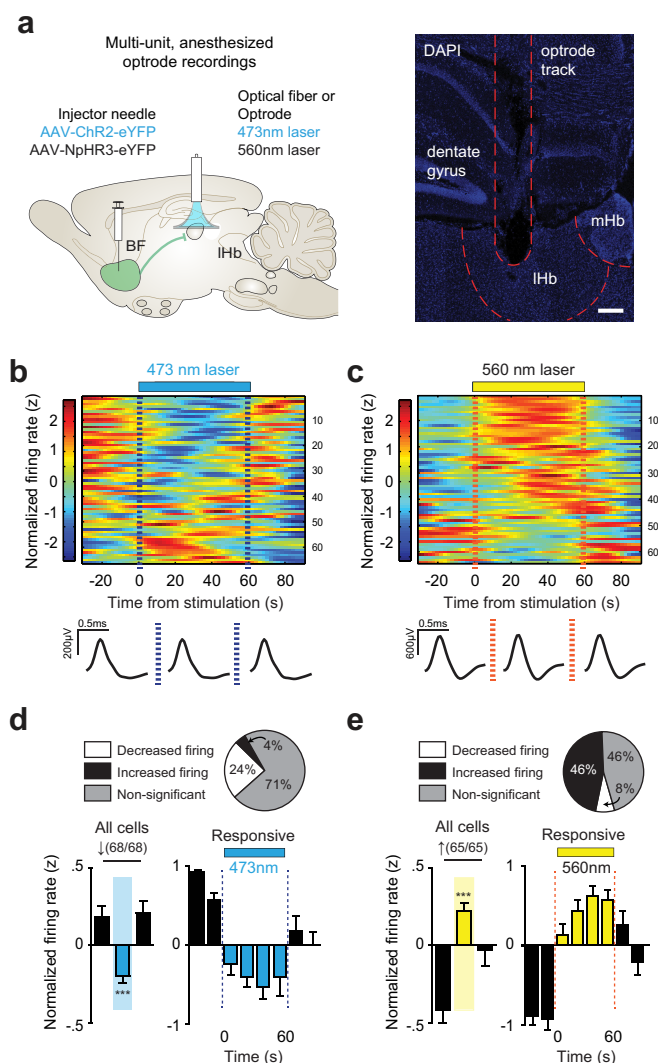
Extended Data Figure 3 | Aggression CPP behaviour. **a**, Experimental schematic of aggression CPP procedure. **b**, **c**, Individual duration spent in the intruder-paired context for AGG ($t_7 = 3.106$, $*P < 0.05$; two-tailed paired t -test, $n = 8$ per group) (**b**) and NON ($t_7 = 2.918$, $*P < 0.05$; two-tailed paired t -test, $n = 8$ per group) (**c**). **d**, Duration spent in the

middle neutral chamber during pre-test and test sessions. **e**, Experimental schematic of sensory CPP procedure. **f**, **g**, Individual duration spent in the intruder-paired context for AGG (**f**) and NON (**g**). **h**, Duration spent in the middle neutral chamber during pre-test and test sessions. Summary data are represented as mean \pm s.e.m.



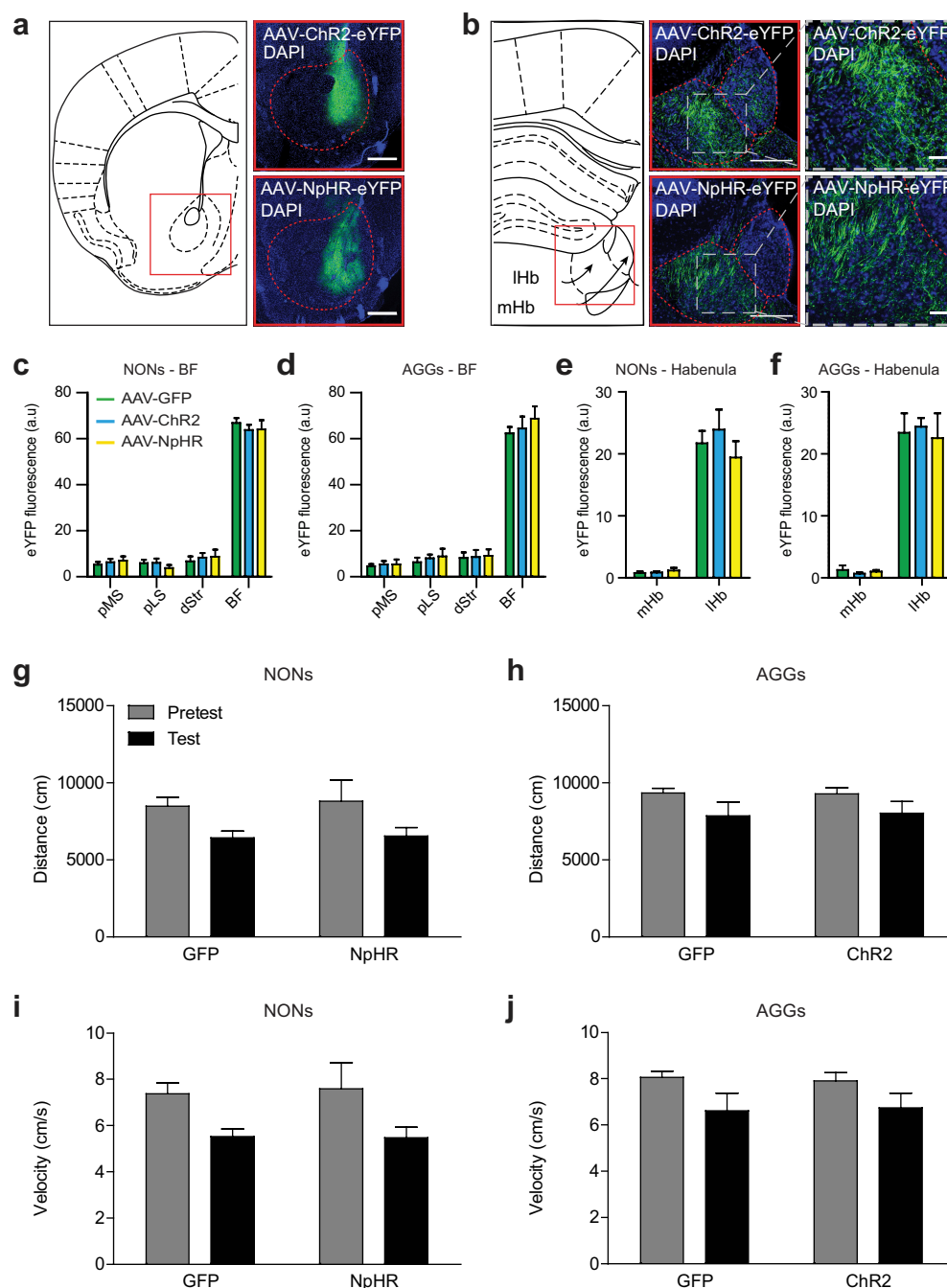
Extended Data Figure 4 | BF-IHb circuit tracing and GABAergic cell-type specificity. **a**, Schematic of viral tracing strategy. **b**, Representative BF viral infection with AAV2-hSyn-eYFP. Scale bar: 500 μ m. **c**, Histological analysis of viral infection with AAV2-hSyn-eYFP ($F_{3,11} = 223.0$, one-way ANOVA $***P < 0.0001$, post-hoc test, $***P < 0.0001$; $n = 3$ mice, 3 slices per mouse) across adjacent anatomical regions. **d**, **e**, Whole-cell electrophysiological recordings (**d**) and representative traces of IHb neurons photostimulated with AAV2-hSyn-ChR2.0 in the absence or

presence of bath-applied GABA_A receptor antagonist gabazine (2 μ M; $F_{2,7} = 220$, one-way ANOVA $P < 0.05$; post-hoc test, $***P < 0.001$, $n = 4$, 2, 2 cells from 2 mice) (**e**). **f**, Optically evoked IPSC response delay ($n = 21$ oIPSC events, 2 mice). **g**, Representative images of eYFP^{BF→IHb} terminal colocalization between vesicular GABA transporter (top), and not vesicular glutamate transporter 1 (bottom). Scale bars: 10 μ m; white arrows indicate colocalization within insets. MS, medial septum; pLS, posterior lateral septum. Summary data are represented as mean \pm s.e.m.



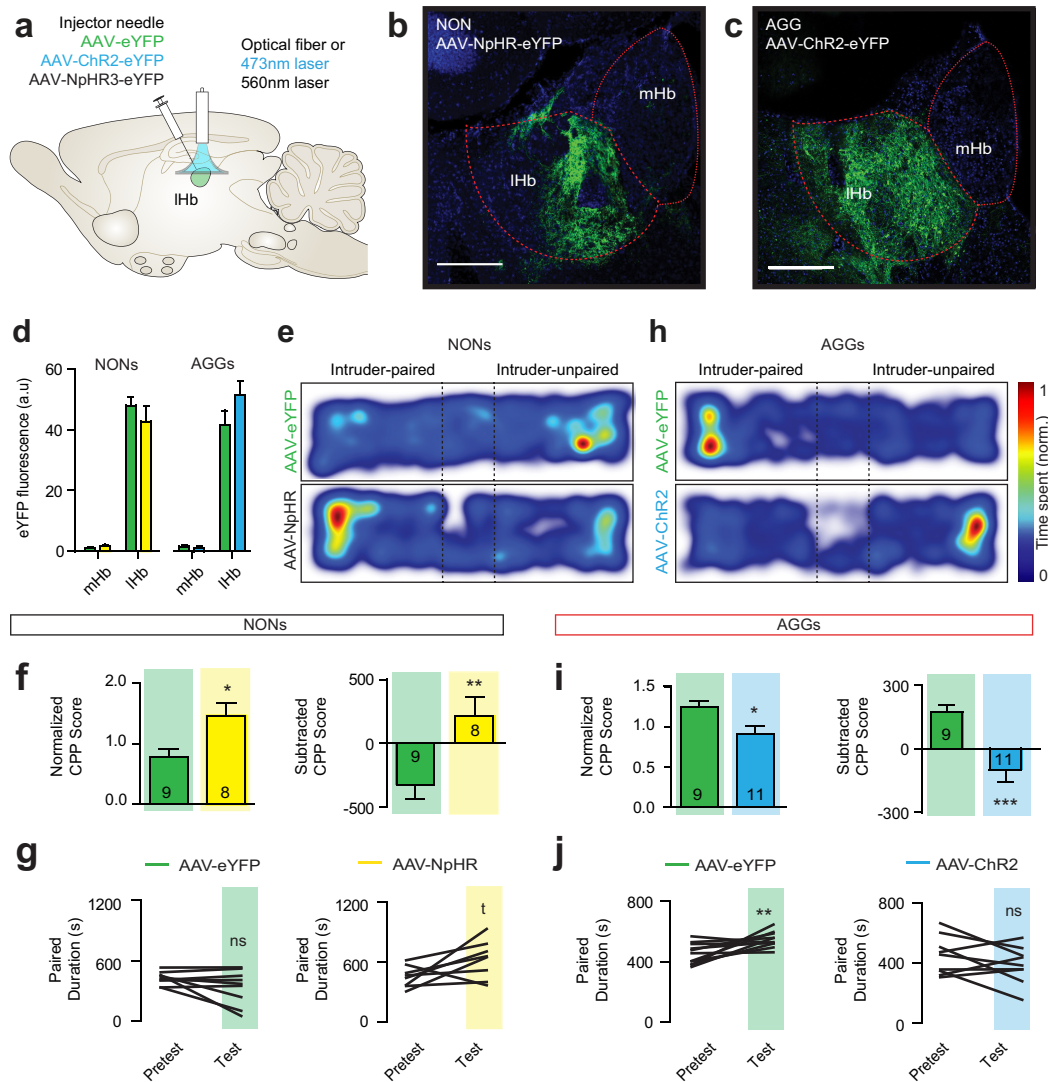
Extended Data Figure 5 | Multiunit anaesthetized optrode recordings.

a, Schematic of *in vivo* anaesthetized multi-unit optrode recording procedure (left) and representative optrode placement in LHb (right; scale bar: 200 μ m). **b**, **c**, Heatmaps of normalized firing rates for LHb neurons in response to BF terminal stimulation with ChR2^{BF→LHb} (**b**) or NpHR3^{BF→LHb} (**c**) and averaged spike wave-form shown below for pre-stimulation, stimulation and post-stimulation epochs. **d**, Percentage of cells by firing response (top) and average normalized LHb firing rate (bottom) after BF–LHb terminal stimulation with ChR2^{BF→LHb} for all identified cells ($F_{2,134} = 8.249$, one-way repeated-measure ANOVA $P < 0.001$; post-hoc test, $*P < 0.05$; $n = 68$ cells from 3 mice) and cells that significantly decreased firing during the stimulation epoch ($F_{7,105} = 8.868$, one-way repeated-measure ANOVA $P < 0.0001$; post-hoc test, $*P < 0.05$; $n = 16/68$ cells from 3 mice). **e**, Percentage of cells by firing response (top) and average normalized LHb firing rate (bottom) after BF–LHb terminal stimulation with NpHR3^{BF→LHb} for all identified cells ($F_{2,128} = 10.32$, one-way repeated-measure ANOVA $P < 0.0001$; post-hoc test, $*P < 0.05$; $n = 65/65$ cells from 3 mice) and cells that significantly increased firing during the stimulation epoch ($F_{7,203} = 17.58$, one-way repeated-measure ANOVA $P < 0.0001$; post-hoc test, $*P < 0.05$; $n = 30/65$ cells from 3 mice). mHb, medial habenula. Summary data are represented as mean \pm s.e.m.



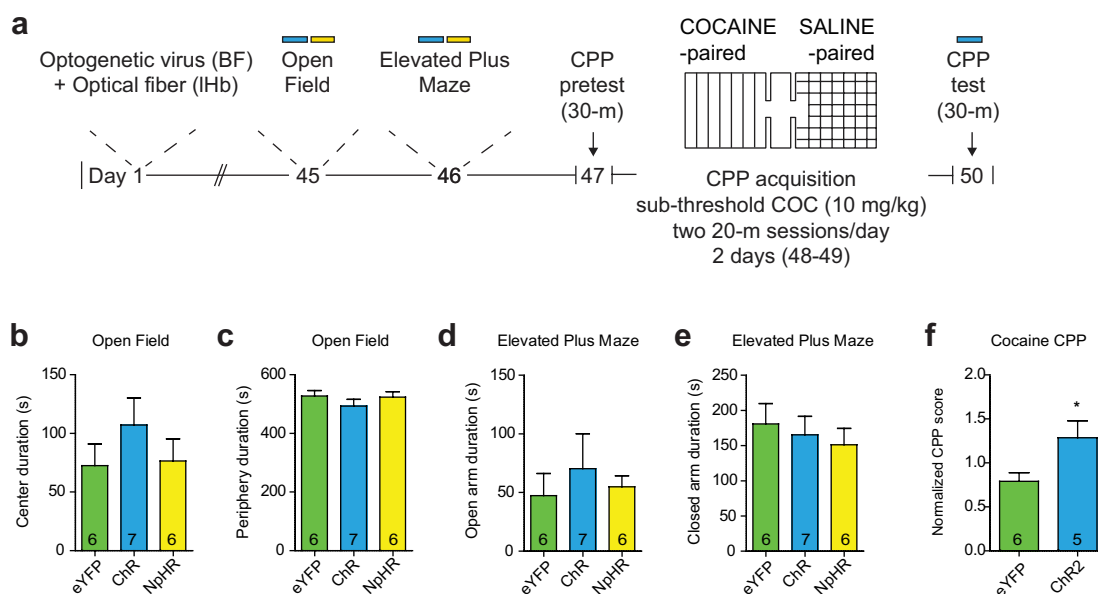
Extended Data Figure 6 | BF-IHb AAV infection and CPP locomotor behaviour. **a**, Schematic of BF coronal slice (left), alongside representative AAV-ChR2-eYFP (top) and AAV-NpHR3.0-eYFP (bottom) infections. Scale bar: 500 μ m. **b**, Schematic of IHb coronal slice (left), alongside representative images of BF terminal infection by AAV-ChR2-eYFP (middle top) and AAV-NpHR3.0-eYFP (middle bottom) within the IHb. Scale bar: 200 μ m. Representative close-ups of terminal regions shown in insets on right. Scale bar: 50 μ m. All representative images counterstained

with DAPI. **c**, **d**, Histological analysis of BF infection in NON (**c**) and AGG (**d**) mice. **e**, **f**, Histological analysis of habenular viral infection in NON (**e**) and AGG mice (**f**). **g**–**j**, Total distance travelled (**g**, **h**) and mean velocity (**i**, **j**) between NON and AGG during the CPP pre-test and test phase. All data are presented as mean \pm s.e.m., and are not significant as determined by two-way ANOVA, $P < 0.05$. dStr, dorsal striatum; mHb, medial habenula; MS, medial septum; pLS, posterior lateral septum.



Extended Data Figure 7 | Direct IHb stimulation bi-directionally modulates aggression reward. **a**, Schematic of viral infection strategy. **b**, **c**, Representative images of IHb cell body infection in NON (**b**) and AGG (**c**). Scale bar: 200 μ m. **d**, Histological analysis of IHb viral infection. **e**, Representative CPP traces of NON. NON::NpHR^{IHb} cell body infection mimics the physiological effect of NON::ChR2^{BF→IHb} terminal stimulation. **f**, Normalized CPP score ($t_{15} = 2.834$, $*P < 0.05$; two-tailed unpaired t -test, $n = 8-9$ per group) and subtracted CPP score ($t_{15} = 3.058$, $**P < 0.01$; two-tailed unpaired t -test, $n = 8-9$ per group) in NON::eYFP^{IHb} and NON::NpHR^{IHb}. **g**, Individual duration spent in the intruder-paired context for NON::eYFP^{IHb} ($t_9 = 0.9129$, $P > 0.05$; two-tailed

paired t -test, $n = 10$ per group) and NON::NpHR^{IHb} ($t_9 = 2.344$, $*P < 0.05$; two-tailed paired t -test, $n = 10$ per group). **h**, Representative CPP traces of AGG::eYFP^{IHb} and AGG::ChR2^{IHb}. **i**, Normalized CPP score ($t_{18} = 2.692$, $*P < 0.05$; two-tailed unpaired t -test, $n = 9-11$ per group) and subtracted CPP score ($t_{18} = 4.203$, $***P < 0.01$; two-tailed unpaired t -test, $n = 9-11$ per group) for the intruder-paired context in AGG::eYFP^{IHb} and AGG::ChR2^{IHb}. **j**, Individual duration spent in the intruder-paired context for AGG::eYFP^{IHb} mice ($t_{10} = 3.212$, $**P < 0.01$; two-tailed paired t -test, $n = 9$ per group) and AGG::ChR2^{IHb} mice ($t_8 = 1.348$, $P < 0.05$; two-tailed paired t -test, $n = 11$ per group). Summary data are represented as mean \pm s.e.m. dStr, dorsal striatum; mHb, medial habenula.



Extended Data Figure 8 | BF-IHb stimulation modulates cocaine CPP. **a**, Experimental timeline of general anxiety and cocaine CPP testing. **b–e**, BF-IHb stimulation during open field testing (**b**, **c**) and elevated plus maze testing (**d**, **e**). **f**, Subthreshold cocaine (10 mg kg⁻¹, intraperitoneal) CPP procedure with BF-IHb stimulation during CPP test ($t_9 = 2.403$, $P < 0.05$; two-tailed unpaired t -test, $n = 5–6$ per group).

Extended Data Table 1 | Stress and anxiety behaviours in AGG and NON

Elevated plus maze

		Duration in (s)		Latency to enter (s)		Distance (cm)	Velocity (cm/s)
		Closed arms	Open arms	Closed arms	Open arms		
NON	Mean	136.53	58.85	13.51	9.06	2412.14	8.16
	SEM	11.71	8.55	9.22	2.37	114.19	0.38
AGG	Mean	137.39	65.07	5.11	5.50	2515.77	8.67
	SEM	6.51	10.48	2.21	1.79	136.55	0.46
P value		0.95	0.65	0.39	0.24	0.57	0.40
n		12	12	12	12	12	12

Open field and locomotion

		Duration in (s)			Latency (s)	Distance (cm)	Velocity (cm/s)
		Center	Middle	Periphery	Center		
NON	Mean	18.80	126.63	373.76	24.35	4909.51	8.77
	SEM	3.28	28.87	58.22	7.78	374.32	0.59
AGG	Mean	15.59	138.70	445.82	31.31	4335.25	8.90
	SEM	1.77	11.63	12.20	7.22	211.61	0.36
P value		0.36	0.82	0.48	0.53	0.73	0.84
n		7-11	7-11	7-11	7-11	7-11	7-11

Forced-swim test

		Duration (s)		Distance (cm)	Velocity (cm/s)	Percent immobile
		Mobile	Immobile			
NON	Mean	260.96	175.76	2041.17	11.40	0.40
	SEM	7.78	8.21	101.09	0.58	0.02
AGG	Mean	275.40	164.19	2076.95	11.57	0.37
	SEM	8.20	8.92	70.68	0.39	0.02
P value		0.22	0.35	0.78	0.81	0.29
n		10	10	10	10	10

Sucrose preference

		Volume consumed (ml)						Sucrose preference		
		Day 1		Day 2		Day 1 + Day 2		Sucrose / (Sucrose + Water)		
		Water	Sucrose	Water	Sucrose	Water	Sucrose	Day 1	Day 2	Day 1 + Day 2
NON	Mean	2.44	5.74	1.54	5.74	3.98	11.48	0.71	0.76	0.75
	SEM	0.65	0.66	0.26	0.69	0.79	0.86	0.07	0.05	0.04
AGG	Mean	2.39	5.15	2.03	7.50	4.42	12.65	0.69	0.78	0.74
	SEM	0.58	0.58	0.59	0.77	0.70	1.05	0.06	0.06	0.04
P value		0.95	0.51	0.47	0.11	0.68	0.41	0.82	0.77	0.84
n		10-11	10-11	10-11	10-11	10-11	10-11	10-11	10-11	10-11

The behavioural data are shown as mean \pm s.e.m. and analysed by unpaired Student's *t*-test. Significance at **P* < 0.05.

Extended Data Table 2 | Social approach behaviours in AGG and NON

Social interaction test

<u>Approach</u>		No target interaction zone (s)		Target interaction zone (s)		No target corner zone (s)		Target corner zone (s)	
		CD-1	C57	CD-1	C57	CD-1	C57	CD-1	C57
NON	Mean	56.62	51.55	77.49	76.42	34.77	32.53	21.54	25.03
	SEM	7.20	7.66	5.25	6.71	5.64	6.23	4.76	2.80
AGG	Mean	55.71	54.17	82.44	79.37	28.24	29.83	20.79	23.20
	SEM	4.67	6.07	6.13	8.91	2.52	5.71	2.84	4.95
P value		0.67	0.80	0.53	0.85	0.56	0.55	0.81	0.82
n		6-7	6-7	6-7	6-7	6-7	6-7	6-7	6-7

<u>Locomotion</u>		No target distance (cm)		Target distance (cm)		No target velocity (cm/s)		Target velocity (cm/s)	
		CD-1	C57	CD-1	C57	CD-1	C57	CD-1	C57
NON	Mean	1501.64	1715.50	1343.64	1447.93	10.01	11.44	8.96	9.65
	SEM	198.09	83.30	153.22	89.24	1.32	0.56	1.02	0.59
AGG	Mean	1544.73	1699.40	1284.39	1474.84	10.31	11.39	8.57	9.87
	SEM	31.04	121.71	41.37	97.81	0.20	0.83	0.28	0.63
P value		0.83	0.92	0.72	0.84	0.82	0.96	0.72	0.81
n		6-7	6-7	6-7	6-7	6-7	6-7	6-7	6-7

Social vs. novel object test

<u>Pretest phase</u>		Duration (s)		CD-1 Target Normalized social score		Subtracted social score	Duration (s)		C57BL6/J Target Normalized social score		Subtracted social score
		Social	Novel	Social	Novel		Social	Novel	Social	Novel	
NON	Mean	114.42	94.84	-	-	19.59	114.91	95.51	-	-	19.40
	SEM	5.30	8.12	-	-	11.78	7.96	7.70	-	-	15.22
AGG	Mean	121.48	97.66	-	-	23.82	106.41	95.02	-	-	11.38
	SEM	6.40	8.34	-	-	12.70	10.22	10.81	-	-	18.66
P value		0.41	0.81	-	-	0.81	0.52	0.97	-	-	0.74
n		7-8	7-8	-	-	7-8	8	8	-	-	8

<u>Test phase</u>		Duration (s)		CD-1 Target Normalized social score		Subtracted social score	Duration (s)		C57BL6/J Target Normalized social score		Subtracted social score
		Social	Novel	Social	Novel		Social	Novel	Social	Novel	
NON	Mean	166.49	93.31	1.50	1.08	73.17	153.47	95.64	1.41	1.04	57.83
	SEM	12.49	10.70	0.16	0.19	22.88	6.41	4.87	0.16	0.09	9.14
AGG	Mean	154.89	100.56	1.30	1.08	54.33	148.77	98.80	1.50	1.13	49.97
	SEM	5.03	7.04	0.09	0.12	10.61	7.60	8.42	0.16	0.14	14.21
P value		0.41	0.59	0.32	0.98	0.49	0.55	0.48	0.85	0.32	0.83
n		7-8	7-8	7-8	7-8	7-8	8	8	8	8	8

The behavioural data are shown as mean \pm s.e.m. and analysed by unpaired Student's *t*-test. Significance at **P* < 0.05. The subtracted social score was derived by subtracting time in the social-paired chamber from the novel object-paired chamber during the test phase. Normalized social score is the ratio of time spent in the chamber of interest (social target or novel object) during the test phase over the pre-test phase.

Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing

Justin Jee^{1,2}, Aviram Rasouly^{1,3}, Ilya Shamovsky¹, Yonatan Akivis¹, Susan R. Steinman¹, Bud Mishra^{2§} & Evgeny Nudler^{1,3§}

In 1943, Luria and Delbrück used a phage-resistance assay to establish spontaneous mutation as a driving force of microbial diversity¹. Mutation rates are still studied using such assays, but these can only be used to examine the small minority of mutations conferring survival in a particular condition. Newer approaches, such as long-term evolution followed by whole-genome sequencing^{2,3}, may be skewed by mutational ‘hot’ or ‘cold’ spots^{3,4}. Both approaches are affected by numerous caveats^{5–7}. Here we devise a method, maximum-depth sequencing (MDS), to detect extremely rare variants in a population of cells through error-corrected, high-throughput sequencing. We directly measure locus-specific mutation rates in *Escherichia coli* and show that they vary across the genome by at least an order of magnitude. Our data suggest that certain types of nucleotide misincorporation occur 10⁴-fold more frequently than the basal rate of mutations, but are repaired *in vivo*. Our data also suggest specific mechanisms of antibiotic-induced mutagenesis, including downregulation of mismatch repair via oxidative stress, transcription–replication conflicts, and, in the case of fluoroquinolones, direct damage to DNA.

De novo mutations in bacteria remain a notoriously difficult target for high-throughput sequencing. Whereas *E. coli* mutate fewer than 1 in 10⁹ bases per generation, high-fidelity polymerases used for library preparation polymerase chain reaction (PCR) cause errors in ~4 out of 10⁶ bases⁸. Illumina machines misread ~1 in 10³ bases⁹. Recent methods, such as barcoding of reads from the same original DNA molecule⁸, have lowered the error rate of sequencing. However, such methods can have low yields¹⁰ and do not address errors introduced by PCR. PCR errors can be overcome using duplex barcoding, which forms a consensus from both strands of a DNA template molecule¹¹. However, even when a small region is targeted¹², duplexing lowers yield even further. The mutational landscape of an RNA virus with mutation rate 10⁴-fold greater than *E. coli* was recently mapped using ‘circle sequencing’. However, this technique is not designed for targeted coverage of a single locus, and its accuracy is limited by sequence read length^{10,13}.

We introduce maximum-depth sequencing (MDS) for detecting extremely rare variants in any region of interest (ROI) in a population of cells (see Methods, Fig. 1a). By synthesizing unique barcodes directly onto the ROI of an original genomic DNA molecule and then copying that molecule using linear amplification, we increase yield (Fig. 1b) and substantially reduce both polymerase and sequencing errors (Fig. 1c). On mock cultures with single-nucleotide mutants spiked in at known concentrations, MDS reliably recovers the expected proportion of mutants at the lowest frequency tested, 10^{–6} (Extended Data Fig. 1). On *in vitro* synthesized DNA templates, MDS reduces the error rate to less than 5 × 10^{–8} per nucleotide sequenced (Fig. 1c, Extended Data Fig. 2). By increasing the number of reads used to call a consensus sequence (*R*), MDS can lower error rate indefinitely, given sufficient coverage (see Methods, error rate of MDS). Application of a second barcode after linear PCR increases accuracy at an even sharper rate and

was used here to demonstrate library preparation efficiency (Extended Data Fig. 2; Supplementary Information: testing sample preparation and PCR efficiency.)

We used MDS to investigate mutation rates in MG1655 *E. coli* grown for ≤ 120 generations. We investigated six ~100-nucleotide ROIs: (1) part of the coding sequence (CDS) of the β subunit of RNA polymerase (*rpoB*), which confers rifampicin resistance when mutated; (2) the 3′ untranslated region (UTR) of *rpoB*; (3) the RNA polymerase ω subunit, *rpoZ*; (4) the CDS of cold-shock response gene *cspE*; (5) the centre of the CDS of penicillin-binding protein gene *mrcA*; and (6) the 3′ end of the CDS of *mrcA*. The last three genes, when knocked out, do not affect cell growth^{14,15}. Whereas *rpoB*, *rpoZ*, and *cspE* are highly transcribed, *mrcA* is one of the least-transcribed genes in *E. coli* under normal conditions¹⁵. All ROIs have balanced AT and CG content, are transcribed on the leading strand, and lack homopolymers >8 nucleotides (nt).

Mutation rates in *E. coli* have been reported from 0.2 × 10^{–10} to 5 × 10^{–10} nt per generation^{3,16,17}. Our calculated rate of mutation in *rpoB* CDS using synonymous substitutions is 4.1 × 10^{–10} nt per generation, comparable to the rate obtained in ref. 17 and at least one long-term evolution experiment using MG1655 (ref. 2). Yet it is also higher than rates calculated by fluctuation assay and long-term evolution on other strains (Fig. 2a, Extended Data Fig. 3). We performed fluctuation assays and recovered a similar spectrum and low rate of mutation to others using such approaches¹⁶. It is likely that the higher rate of mutation in *rpoB* obtained with MDS indicates a rate uninfluenced by negative selection, phenotypic lag, or imperfect plating efficiency⁵.

Mutation rate in nonessential *rpoZ*, and *cspE*, as well as *rpoB* UTR, is only slightly higher than that in essential *rpoB* CDS, but our calculated rate of mutation in the middle of *mrcA* is 3.5 × 10^{–9} nt per generation, an order of magnitude higher than the observed rate in *rpoB* CDS and significantly higher than the rates of mutation in all other ROIs (*P* < 0.001 by ANOVA). The 3′ end of *mrcA* also has a higher rate of mutation than all other ROIs considered except for the middle of *mrcA*, suggesting spatial clustering of mutation rates. Comparison of genomes from several *E. coli* strains has suggested that clustered, highly transcribed genes are protected from mutation by an unknown mechanism⁴, a finding that has since been challenged^{3,18}. Our results demonstrate that at least one gene with low transcription rate has significantly higher mutation rate than three others with high transcription rate.

The mutational spectrum from MDS matches that found in long-term sequencing experiments, with transition mutations favoured over transversions (Fig. 3a, Extended Data Figs 4, 5a). We also note an unexpected high frequency of C→A substitutions. These do not appear to be lasting mutations, as complementary G→T substitutions emerged with less than 0.1-fold frequency. A similar effect was found to a lesser extent for G→A and C→T substitutions. Increasing *R* did not significantly reduce these high substitution frequencies (Fig. 3b, Supplementary Information: model of damaged base pairs),

¹Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA. ²Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA. ³Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA.

§These authors jointly supervised this work.

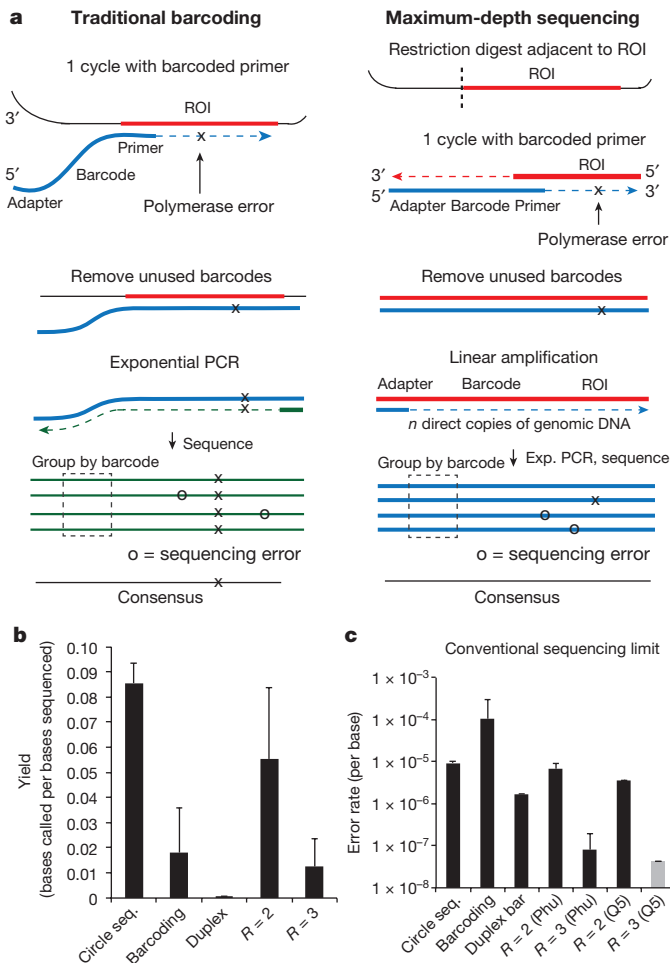


Figure 1 | Overview of MDS. a, Comparison of traditional barcoding protocol with MDS (see Methods for details). Note an additional barcode can be attached after linear amplification to further increase accuracy (see Extended Data Fig. 2). **b**, Mean yield of various methods, in consensus nucleotides called per nucleotides sequenced. Results from our study are boxed. **c**, Mean error rate of various methods when applied to DNA synthesized *in vitro*, in frequency of miscalled bases (log₁₀ scale). Error rates from our study are given using both Phusion (Phu) and Q5 polymerase. Q5, $R=3$. Analysis of 1,685,502 consensus nucleotides yielded no errors. The value shown is extrapolated from the Q5 error rate and expected reduction given $R=3$. Yield and error rate from previous methods are from ref. 10. MDS experiments were performed in quadruplicate. Error bars are s.d.

suggesting that the majority of *in vivo* C→A substitutions are not due to damaged nucleotides. We found that *in vitro* templates synthesized with 8-oxoguanine (8-oxoG) resulted in low C→A substitution rates (Extended Data Fig. 3c), and treatment of *in vivo* DNA with formamidopyrimidine DNA glycosylase (FPG) did not change the observed substitution frequency (Extended Data Fig. 3c), further confirming that these C→A substitutions are probably not due to 8-oxoG. It is possible that As, or ribonucleotide As, are misincorporated into the genome at C sites *in vivo*. We found that neighbouring Cs are predictive of a higher frequency of C→A substitutions, suggesting that these transient substitutions cluster spatially along the genome, unlike polymerase or sequencing errors (Fig. 3c, Extended Data Figs 3b, 4, 5b).

In vivo, these misincorporations must be reversed before genome replication. However, our observations represent a snapshot of this dynamic process before repair can occur. Although these events would be invisible to conventional methods, the frequency of these substitutions, at $\sim 10^{-5}$ per nucleotide, is over 10^4 times more frequent than the true rate of mutation.

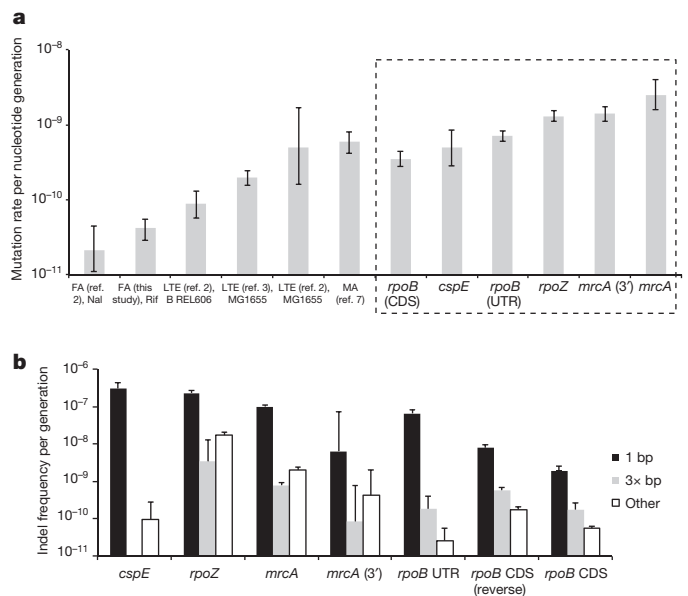


Figure 2 | Substitution rates and indel frequencies. a, Comparison of mutation rates calculated from fluctuation assays (FA) using either rifampicin (Rif) or nalidixic acid (Nal), long-term evolution (LTE), and mutation accumulation (MA). Rates calculated using MDS are boxed. Note that number of generations is calculated according to population doubling time in refs 2 and 3 (see Supplementary Information: generation time models). **b**, Frequency of indel mutations recovered at $t=120$ generations. Values are normalized for possible indel lengths considered in each category. Experiments are biological quadruplicates. All error bars are 95% confidence intervals (CI).

To clarify which substitutions are transient rather than involved in ‘true’ mutation, we analysed DNA from bacteria collected after ≤ 20 generations, a short enough time period to expect few true mutations, given our sample size (Fig. 3a). We observed enrichment for most types of substitutions in our ≤ 120 generation trial over our ≤ 20 generation control, as would be expected from true mutations. However, C→A, A→G, and C→T substitutions occur in comparable frequency in the 20 and 120 generation trials, suggesting these substitutions reflect a continual process of base misincorporation and repair. We did not include these abundant A and T substitutions in our calculation of mutation rates. However, these findings suggest that the mechanism underlying the increase of AT content in *E. coli* grown for long periods² is a dynamic process of misincorporation and repair.

We calculated short (≤ 12 base pairs (bp)) indel rates in *mrcA*, *rpoB* UTR, *rpoZ*, and *cspE* ROIs (Fig. 2b). Indel rate varies widely by position and size. As might be expected¹⁹, 100% of the observed 1-bp indels occur at a site adjacent to a homopolymer. The frequency of 1-bp indels also increases with homopolymer length, potentially explaining why *cspE*, with an 8-bp T homopolymer, has the highest 1-bp indel rate. Longer indels are not localized to homopolymers and are positively correlated with substitution rates across all ROIs (Extended Data Fig. 6), supporting previous work suggesting that indels and substitutions spatially cluster in comparisons of genomes from divergent bacterial species²⁰. In all ROIs, deletions were detected at >10 -fold frequency of insertions.

Single nucleotide indels and longer frameshifting mutations were also observed in *rpoB* CDS, albeit at low frequency, even though such mutations should be deleterious. As expected, the rate of in-frame indels was higher than the rate of frameshift indels of >1 -bp length (Fig. 2b). Because of the low rate of indel errors from *in vitro* polymerases used here⁸, it is plausible that the observed frameshift mutations are from inviable bacteria, as DNA from such cells may still enter our protocol. The recovery of frameshift indels, as well as the nonsignificant difference between rates of synonymous and nonsynonymous

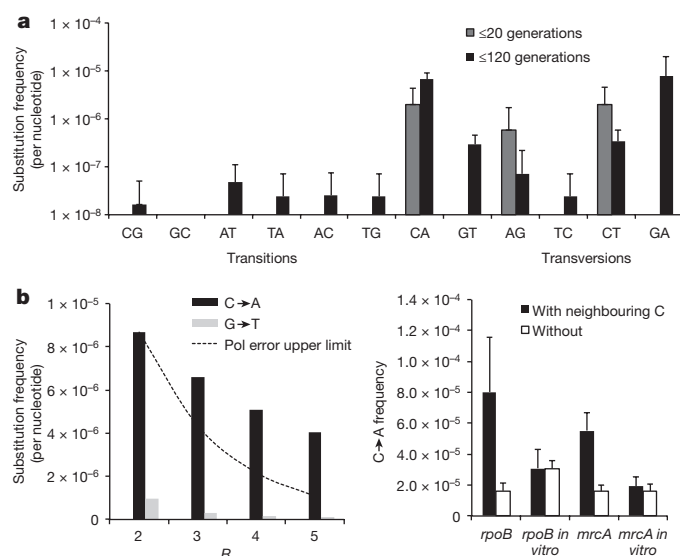


Figure 3 | Substitution spectra. **a**, Frequency of base substitutions recovered in our sequencing protocol at $t = 20$ generations and $t = 120$ generations in *rpoB* CDS. Values are not normalized by number of generations and thus are true frequencies, not mutation rates. Experiments are biological quadruplicates. **b**, The high frequency of C→A substitutions is consistent even as R increases. If these substitutions were polymerase errors due to damaged nucleotides, they should decline with increasing R faster than the line representing a model in which the polymerase makes C→A errors with 50% frequency for a subpopulation of DNA molecules (see Supplementary Information: model of damaged base pairs). **c**, C→A substitutions *in vivo* cluster in nucleotides with at least two neighbouring Cs within a 2-bp radius ($P < 0.01$ by t -test), unlike polymerase errors. Error bars are 95% CI upper bound.

substitutions in *rpoB* CDS (Supplementary Table 3), demonstrate that selection in our protocol is minimal.

Exposing *E. coli* to sub-inhibitory doses of multiple classes of antibiotics increases the rate at which bacteria acquire resistance to rifampicin. Whether this increase is caused by nucleotide oxidation^{21,22}, downregulation of mismatch repair²³, or an unrelated pathway²⁴, has become a topic of interest. We investigated the effect of sub-inhibitory doses of ampicillin and norfloxacin—a β lactam and fluoroquinolone respectively—on mutation rate using MDS of *rpoB* CDS and *mrcA*, as well as detailed fluctuation assays^{16,25} (Fig. 4a). Addition of ampicillin increased the rate of transition mutations in *rpoB*, a signature indicative of downregulated mismatch repair³. In cells overexpressing catalase, basal mutation rate decreased by a factor of 8 (Fig. 4b), indicating that background oxidation contributes significantly to the basal mutation rate under non-stressed conditions. Addition of ampicillin during catalase overexpression did not increase this low rate (Fig. 4b). Overexpression of a catalase with inactivating point mutation H106Y did not confer similar mutagenic protection (Extended Data Fig. 7). These results together support a model in which ampicillin causes oxidative stress²¹, which acts upstream of downregulation of mismatch repair²³ to increase mutation rate. Consistently, cells grown in anaerobic conditions did not display an increase in transition rate when challenged with ampicillin (Extended Data Fig. 8a). The same was true in aerobic conditions if mismatch repair gene *mutS* was knocked out (Extended Data Fig. 8b; see Supplementary Information for further discussion).

In contrast, exposure to norfloxacin increased the rate of >1-bp indel formation in both *mrcA* and *rpoB* (Fig. 4a). Norfloxacin inhibits DNA gyrase and can cause double-strand breaks in DNA²⁶. This physical interaction thus directly causes antibiotic-induced mutagenesis in norfloxacin-treated cells.

There is debate as to whether highly transcribed genes in bacteria have a higher^{18,27} or lower⁴ mutation rate than other genes. Our analysis in *E. coli* shows that *mrcA* has a higher basal rate of mutation than

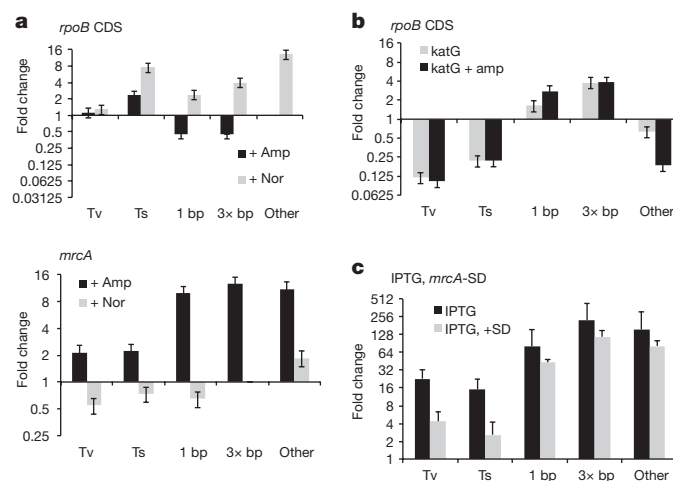


Figure 4 | Relationships between mutation rates and physiologic conditions. **a**, Fold change in transversion (Tv), transition (Ts), and indel rate in response to ampicillin or norfloxacin according to MDS (for fluctuation assay results and raw substitution rates see Extended Data Fig. 9). **b**, Fold change in mutation rate in a strain overexpressing catalase (*katG*). **c**, Fold change in mutation rate of *mrcA* in response to induction via IPTG promoter. Experiments are biological quadruplicates. Error bars are 95% CI. SD, Shine–Dalgarno sequence.

more highly transcribed genes. Yet interestingly, addition of ampicillin increased transversions and indel formation in *mrcA*, but not in *rpoB* CDS (Fig. 4a). It is known that *mrcA* undergoes mild induction upon addition of ampicillin²⁸. To study the effect of transcription on mutagenesis further, we created a strain in which a chromosomal copy of *mrcA* is regulated by an isopropyl β -D-1-thiogalactopyranoside (IPTG) promoter. Induction of *mrcA* transcription increased the frequency of all classes of *mrcA* substitution and indel ~8-fold more than when wild-type cells were exposed to ampicillin (Fig. 4c). These results suggest that although, in basal conditions, cells may have a means of protecting the most highly transcribed genes, co-directional collisions between transcription and replication machinery, which can cause double-strand breaks²⁹, are themselves mutagenic. Induction itself may thus be an important mechanism of stress-induced mutagenesis³⁰.

The low translation rate of *mrcA*, coupled with our finding that *rpoB* UTR has a higher rate of mutation than the CDS, suggests that translation may be protective for highly transcribed genes. We constructed an additional strain in which IPTG-regulated *mrcA* has a canonical Shine–Dalgarno sequence and start codon, rather than its low-translation endogenous sequence. Increasing translation decreased substitution rate in the IPTG-induced state by a factor of 50% and a factor of 75% when high-frequency (C→A, for example) substitutions are excluded (Fig. 4c). Although translation does not lower the *mrcA* mutation rate to *rpoB* levels, it probably contributes to protection of highly transcribed genes (Supplementary Information: relationship between transcription, translation, and mutation rate).

Straightforward extensions to MDS would allow for analysis of many ROIs simultaneously and assembly of longer ROIs (Supplementary Information: MDS protocol). MDS may also be useful in detection of genetic abnormalities in cell-free DNA due to foetal mutations or cancer.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 January; accepted 5 May 2016.

Published online 22 June 2016.

- Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
- Wielgoss, S. *et al.* Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 (Bethesda)* **1**, 183–186 (2011).

3. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783 (2012).
4. Martincorena, I., Seshasayee, A. S. N. & Luscombe, N. M. Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature* **485**, 95–98 (2012).
5. Lenski, R. E., Slatkin, M. & Ayala, F. J. Mutation and selection in bacterial populations: alternatives to the hypothesis of directed mutation. *Proc. Natl Acad. Sci. USA* **86**, 2775–2778 (1989).
6. Wielgoss, S. *et al.* Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl Acad. Sci. USA* **110**, 222–227 (2013).
7. Drake, J. W. Contrasting mutation rates from specific-locus and long-term mutation-accumulation procedures. *G3 (Bethesda)* **2**, 483–485 (2012).
8. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108**, 9530–9535 (2011).
9. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* **12**, R112 (2011).
10. Lou, D. I. *et al.* High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl Acad. Sci. USA* **110**, 19872–19877 (2013).
11. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. USA* **109**, 14508–14513 (2012).
12. Schmitt, M. W. *et al.* Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat. Methods* **12**, 423–425 (2015).
13. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).
14. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
15. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
16. Garibyan, L. *et al.* Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair (Amst.)* **2**, 593–608 (2003).
17. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164 (1991).
18. Chen, X. & Zhang, J. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol. Biol. Evol.* **30**, 1559–1562 (2013).
19. McDonald, M. J., Wang, W.-C., Huang, H.-D. & Leu, J.-Y. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol.* **9**, e1000622 (2011).
20. Zhu, L., Wang, Q., Tang, P., Araki, H. & Tian, D. Genomewide association between insertions/deletions and the nucleotide diversity in bacteria. *Mol. Biol. Evol.* **26**, 2353–2361 (2009).
21. Kohanski, M. A., DePristo, M. A. & Collins, J. J. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol. Cell* **37**, 311–320 (2010).
22. Dwyer, D. J. *et al.* Antibiotics induce redox-related physiological alterations as part of their lethality. *Proc. Natl Acad. Sci. USA* **111**, E2100–E2109 (2014).
23. Gutierrez, A. *et al.* β -Lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat. Commun.* **4**, 1610 (2013).
24. Liu, Y. & Imlay, J. A. Cell death from antibiotics without the involvement of reactive oxygen species. *Science* **339**, 1210–1213 (2013).
25. Lang, G. I. & Murray, A. W. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* **178**, 67–82 (2008).
26. Pohlhaus, J. R. & Kreuzer, K. N. Norfloxacin-induced DNA gyrase cleavage complexes block *Escherichia coli* replication forks, causing double-stranded breaks *in vivo*. *Mol. Microbiol.* **56**, 1416–1429 (2005).
27. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **10**, 449–458 (2012).
28. Sangurdekar, D. P., Sreenc, F. & Khodursky, A. B. A classification based framework for quantitative description of large-scale microarray data. *Genome Biol.* **7**, R32 (2006).
29. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533–543 (2011).
30. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat. Rev. Genet.* **2**, 504–515 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Heguy and the NYU Genome Technology Center, which is partially supported by the Cancer Center Support Grant, P30CA016087, at the Laura and Isaac Perlmutter Cancer Center. This work used computing resources at the High Performance Computing Facility of the Center for Health Informatics and Bioinformatics at the NYU Langone Medical Center. We thank D. Dwyer and K. Shankarling for materials, and T. Artemyev for his contribution. This work was supported by NIH grant R01GM107329 and HHMI (E.N.) and NCI PSOC grant U54 CA193313 (B.M.). J.J. was supported by the NYU Medical Scientist Training Program and a National Defense Science and Engineering Graduate Fellowship.

Author Contributions J.J. and I.S. designed the MDS protocols. J.J., A.R., and E.N. designed the biological experiments. J.J., A.R., and Y.A. performed the experiments. J.J., B.M., S.S., and I.S. performed the data analysis. J.J. and E.N. wrote the manuscript with input from all co-authors. B.M. and E.N. supervised the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.N. (evgeny.nudler@nyumc.org) or B.M. (mishra@nyu.edu).

Reviewer Information *Nature* thanks N. Luscombe, I. Martincorena, J. Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size.

Maximum-depth sequencing. First, genomic DNA is treated with a restriction enzyme, which cleaves at the 3' end of the ROI. A single PCR cycle is performed with barcoded primers annealing to the 3' end of the ROI. Because of the exposed 3' site on the genomic DNA molecule left by the restriction enzyme, the genomic DNA molecule acts as a 'primer', causing the barcode and an adaptor to be synthesized onto the end of the ROI. This synthesis effectively attaches the barcode to the original genomic DNA molecule. Unused barcoded primers are removed, and N cycles of linear amplification are performed using only primers to the forward adaptor sequence. This step is important for screening polymerase errors. The polymerase may make an error in any single round of synthesis, increasing the probability of generating a faulty read by N , but by copying the same original DNA molecule multiple times, the probability of recovering a defective copy after analysis is reduced by a factor of N^R , where R is the number of independent reads used to build a consensus sequence. Thus the total error reduction is $1/N^{R-1}$ fold (see below). In this study, typically $N = 12$ and $R \geq 3$, although the empiric value of N after accounting for inefficiencies in PCR is somewhat lower (see Extended Data Fig. 2 and Supplementary Information: testing sample preparation and PCR efficiency). Note that one could also attach a second barcode to each read after linear amplification but before exponential amplification—doing so could allow one to reduce the error rate even further by ensuring multiple reads from the linear amplification step are used in the analysis. By targeting a ROI, we can also use paired-end sequencing to increase yield. Detailed error rate spectra for both Phusion and Q5 polymerase are measured and reported in Supplementary Table 1. It should be noted that when $R > 2$, MDS errors such as those shown in Fig. 1 are derived almost entirely from transition substitutions typical of PCR polymerases, and that for other kinds of substitutions, error rate is virtually nonexistent. In MDS, each read represents additional $1 \times$ coverage of the ROI. Thus, MDS can achieve $\sim 10^9$ -fold coverage using an Illumina HiSeq machine. For details on the specific enzymes, primers, and PCR conditions used in this study see Supplementary Information: MDS protocol. For details on consensus base calling, see Supplementary Information: analysis.

Error rate of MDS. Sources of error include damaged DNA during extraction, polymerase errors during PCR, and sequencing errors. Because our goal is to identify rare mutants, we consider error as the rate of false positives, which affect mutant frequency to a much larger extent than false negatives⁸.

If the probability of a single nucleotide X being misread as Y owing to polymerase error is $P_{\text{pol},X \rightarrow Y}$ and the rate of the corresponding sequencing error is $P_{\text{seq},X \rightarrow Y}$, then the probability that X will be read as Y owing to either source of error in a standard sequencing protocol is

$$P_{X \rightarrow Y} = P_{\text{pol},X \rightarrow Y} + P_{\text{seq},X \rightarrow Y} \quad (1)$$

As discussed briefly in the main text, in our assay, the total polymerase error rate $E_{\text{pol},X \rightarrow Y}$ can be derived as follows (for visual aid, see Extended Data Fig. 10). For convenience, $P_{\text{pol},X \rightarrow Y}$ will hereafter be referred to as p . After exponential PCR, there are N pools of reads, each derived from one of the original linear amplification steps. The probability of having k pools derive from an original polymerase error is binomially distributed. Furthermore, because $Np \ll 1$, the distribution is Poisson.

$$\binom{N}{k} p^k (1-p)^{N-k} \approx \frac{(Np)^k}{k!} e^{-Np} \quad (2)$$

The probability of a false positive is the probability that all R reads used to form a consensus came from one of the k 'error' pools

$$\sum_{k=1}^N \binom{k}{R} \frac{(Np)^k}{k!} e^{-Np} = \frac{1}{N^R} \sum_{k=1}^N k^R \frac{(Np)^k}{k!} e^{-Np} = \frac{1}{N^R} M_R = \frac{B_R(Np)}{N^R} \quad (3)$$

Where M_R is the R th moment of the Poisson distribution in equation (2) and B_R is the R th Bell polynomial. Because $Np \ll 1$, an upper bound on this error formula can be written as follows:

$$E_{\text{pol},X \rightarrow Y} = \frac{B_R(Np)}{N^R} < \frac{p B_R(1)}{N^{R-1}} < \frac{p}{N^{R-1}} \left(\frac{0.792R}{\ln(R+1)} \right)^R \quad (4)$$

Where the upper bound of the R th Bell number $B_R(1)$ is from ref. 31. These bounds will decrease rapidly as R increases, given that $R \leq N$.

We note that in practice, the probability that the same error would emerge in $k > 1$ reads produced by the linear amplification step is $\sim 10^{-12}$, so low that the expected number of such multi-errors for all the nucleotides sequenced in this study is < 1 . With this in mind, it is possible to simplify equation (4) so that the

Bell number term is a non-contributor to the total error. Under this assumption, the probability of false positive is

$$E_{\text{pol},X \rightarrow Y} \approx \frac{P_{\text{pol},X \rightarrow Y}}{N^{R-1}} \quad (5)$$

The above formula for $E_{\text{pol},X \rightarrow Y}$ only takes into account errors introduced during linear amplification. However, the maximum error that could be contributed during a subsequent round of doubling, or exponential, PCR (D) can be found by substituting $N2^D$ for N in the equation above. The sum of all possible errors from all rounds of PCR would thus be

$$E_{\text{total pol},X \rightarrow Y} \approx \sum_D \frac{P_{\text{pol},X \rightarrow Y}}{(N2^D)^{R-1}} \quad (6)$$

For $R = 2$, this will be a geometric series with sum no greater than $2E_{\text{pol},X \rightarrow Y}$. For $R > 2$, the sum will be closer to $E_{\text{pol},X \rightarrow Y}$.

The error rate of sequencing after forming a barcode, as discussed thoroughly in other texts^{8,10} is the probability that the same error happens \tilde{R} times

$$E_{\text{seq},X \rightarrow Y} = (P_{\text{seq},X \rightarrow Y})^{\tilde{R}} \quad (7)$$

Where \tilde{R} is the number of 'not necessarily independent' reads used to form a consensus (that is, overlapping paired-end sequences of the same read are included). If single-end sequencing is used, $\tilde{R} = R$. If paired-end sequencing is used, a maximum of $\tilde{R} = 2R$ not necessarily independent reads are used.

Alternatively, one could estimate $E_{\text{seq},X \rightarrow Y}$ based on the sum of the quality scores of the \tilde{R} reads contributing to the consensus, but in practice we find this to be unnecessary because sequencing errors are not the major contributor to overall error when $R > 2$.

The total error rate for any given nucleotide position is the sum of all $E_{X \rightarrow Y}$, $X \neq Y$, for a given X . The values reported in the main text and Fig. 1c are total error. Raw polymerase and sequencing error rates⁹ are shown in Supplementary Table 1. Note that this model is also the basis for the damaged base-pair analysis presented in Fig. 3b and the Supplementary Information.

Growth and mutation rate analysis. *E. coli* were streaked onto Luria-Bertani (LB) agar from freezer stocks and grown at 30 °C for 24 h. According to plating and colony-forming unit (c.f.u.) counting, the average number of cells in such colonies is 3×10^8 (thus the number of generations is $\ln(3 \times 10^8) = 19.5$). Bacteria from a single colony were used to inoculate a small liquid culture (1 ml LB broth in a round-bottom tube). For the purposes of generation counting, it is assumed that after the transition to growing in liquid, growth occurs for only ~ 3 generations. The culture was grown in a 37 °C shaker to allow for the transition to growth in broth for 12 h, after which a measurable optical density could be reliably detected.

4 μ l ($\sim 10^7$ bacteria) were transferred to a fresh 100 ml LB liquid culture (in a 250 ml Erlenmeyer flask). Liquid cultures were grown for 24 h on a 37 °C shaker, to a density of 2.5×10^9 bacteria according to cell counts (for a total of 2.5×10^{11} bacteria). This process was repeated 9 times. The average number of generations a bacterium would have grown in each liquid culture is

$$\ln \left(\frac{2.5 \times 10^{11}}{10^7} \right) = 10.1 \text{ generations} \quad (8)$$

Thus the average total number of generations g is $19.5 + 3 + 9 \times 10.1 = 113$.

In addition to the large passage size, we stop passaging hundreds of generations before selective sweeps are expected to occur³² and, importantly, long before selection for a hyper-mutating strain might be expected⁶. We also performed simulations to test the effects the probability that any two bacteria have the same founder given expected conditions of passage size (see Supplementary Information: calculation of mutation rate).

Mutation rates μ in our assay are chosen to maximize the likelihood of recovering the mean mutant frequency f for substitutions of a given type $X \rightarrow Y$, which we find are well approximated by a Poisson process over a certain number of generations (in this case 113).

$$\mu_{X \rightarrow Y} = \frac{f(\bar{Y}) - E_{X \rightarrow Y}}{g} \quad (9)$$

More precisely, the frequency is defined as the number of barcode groups with a given mutation divided by the total number of barcode groups under consideration. For example, if $R \geq 3$, then $f(\bar{Y})$ is the number of read families of size ≥ 3 with mutation Y divided by the total number of read families of size ≥ 3 . Mutation rates given in Fig. 2 are computed from the average across all X of $\sum_{Y \neq X} \mu_{X \rightarrow Y}$, with $C \rightarrow A$, $G \rightarrow T$, $C \rightarrow T$, and $G \rightarrow A$ substitutions, excluded for aforementioned reasons. In their place, a correction term (the average transversion or transition

rate based on all other substitutions) is used so that the mutation rate is not systematically underestimated.

Four biological replicates of each condition were grown. All liquid cultures, including the small founding culture, had the possible addition of $1 \mu\text{l ml}^{-1}$ ampicillin or 15 ng ml^{-1} norfloxacin. Cultures for the short-term growth assay and mock culture were grown similarly except without passaging (Supplementary Information: mock culture and short growth assay).

Strains. MG1655 *E. coli* were used as wild-type cells for all experiments. The IPTG-regulated *mrcA* strain MG1655 and IPTG-regulated strain with modified Shine–Dalgarno were recombineered according to ref. 33. For details, see Supplementary Information: strains. For details on the catalase overexpression mutant and inactive H106Y catalase overexpression mutant see ref. 22. In the *mutS* knockout strain, MG1655 *mutS* was replaced with a kanamycin resistance cassette.

Preparation of DNA samples

Genomic DNA. Up to 5 ml of bacterial liquid culture were spun down (see later section for specific growth conditions). Cells were resuspended in 500 μl Tris-EDTA buffer (pH 7.5), and 1,000 units of Ready-Lyse (Epicentre) added; before incubation at room temperature for 1 h and freezing at -80°C overnight. Genomic DNA extraction was performed using Qiagen genomic tip (100G), but without lysozyme and quantified using Nanodrop.

In vitro DNA. Single-stranded oligonucleotides with sequences corresponding to MG1655 *rpoB* at position 1511–1632 and *mrcA* at 1258–1379 were ordered from IDT and resuspended in deionized water. These oligonucleotides were used directly as input to the Extreme-depth sequencing protocol above for calculation of error rate in Fig. 1 and the ‘negative control’ rows in Supplementary Table 1. Note, as expected from quality control reports from IDT, we found a large number of indels in the *in vitro* synthesized templates ($\sim 1\%$ of molecules had some type of indel). However, the fact that we recovered a low substitution rate could be used to confirm the chemical purity of the mononucleotide pools used for synthesis by IDT.

Separately, 10 ng of the same DNA oligonucleotides were used as templates for a standard 20-cycle exponential PCR reaction with only the ROI-annealing component of the forward and reverse primers above using either Q5 or Phusion polymerase. The amplified DNA was used as input into the MDS protocol and used to calculate the intrinsic substitution error rate of those two polymerases as reported in Supplementary Table 1.

Sequencing depth. On average, we divide single HiSeq Rapid Runs of $\sim 240 \text{ M}$ reads into four different ‘conditions’, each corresponding to a particular ROI from

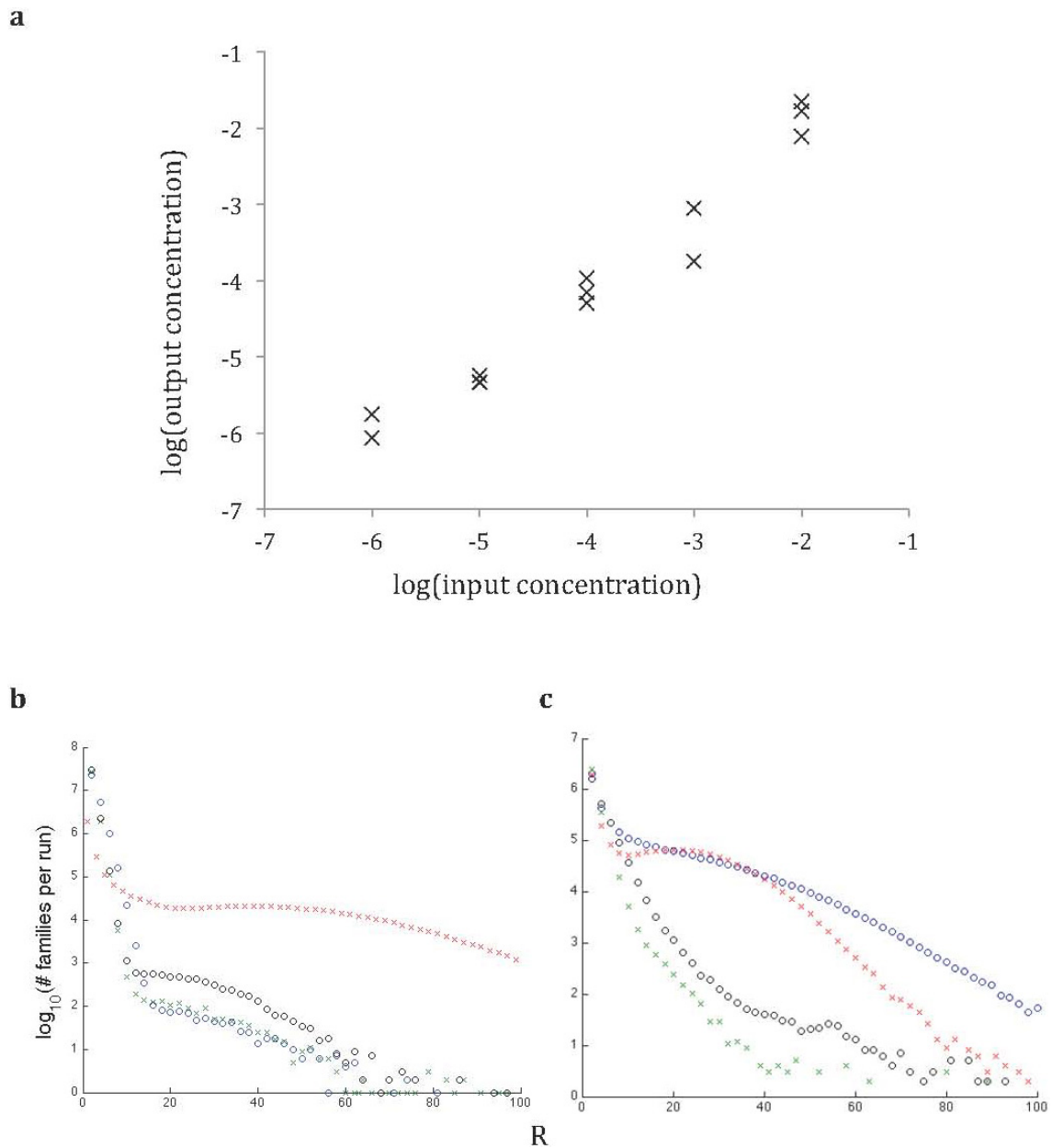
bacteria grown under a certain condition. The $\sim 60 \text{ M}$ reads of each condition are further subdivided in order to process triplicate or quadruplicate trials.

We recover $\sim 2.5 \text{ M}$ total barcode ‘families’ for each condition using our threshold of $R \geq 3$ (for the purposes of calculating total yield, we divide by 2 since each read is pair-end sequenced). We examine $\sim 100 \text{ bp}$ per ROI, thus providing a significant pool from which to observe mutations. There is an interesting level of variability across quadruplicates, likely due to stochastic variation when combining and purifying DNA samples and in binding to the HiSeq flowcell itself (Extended Data Fig. 1b, c). Note that when mutation frequencies are averaged over multiple trials, each trial is weighted according to its relative representation in terms of number of families.

Fluctuation assays. Fluctuation assays were carried out as in ref. 16. We picked single colonies of *E. coli* as above and grew them in 1 ml Luria-Bertani (LB) broth overnight. $0.1 \mu\text{l}$ from this starter culture was used to inoculate 25 separate trial cultures. Each trial culture was grown (in a 37°C shaker) to either an optical density (OD_{600}) of 0.3 (for exponential growth trials) in 2 ml LB broth or for 24 h (for saturation) in 0.2 ml LB broth and plated cultures on Petri dishes containing LB agar with 100 mg ml^{-1} rifampicin. Colonies were grown for 48 h in 30°C and c.f.u. were counted. The *rpoB* region conferring rifampicin resistance was sequenced and used to compute the mutational profiles in Fig. 4. Number of bacteria per culture was calculated by serial dilution, plating on LB agar, and counting c.f.u. after 48 h growth in 37°C . Mutation rates and 95% CIs were computed using the Ma-Sandri-Sarkar method³⁴ as implemented in ref. 25. Broth was possibly supplemented $1 \mu\text{l ml}^{-1}$ ampicillin, 15 ng ml^{-1} norfloxacin, or 250 ng ml^{-1} gentamycin. LB broth was placed in an LS-580 anaerobe chamber (Anaerobe Systems) overnight to yield anaerobic media.

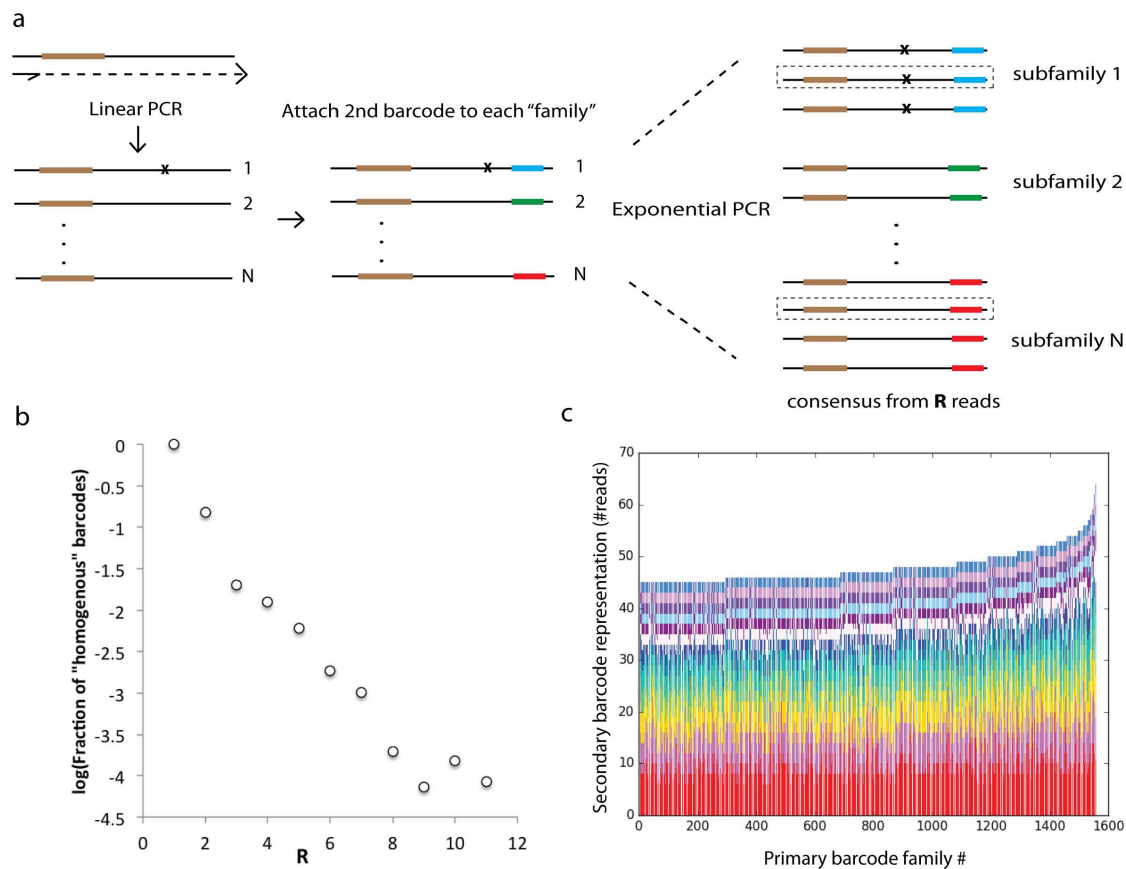
Availability. Raw sequence data are available from Sequence Read Archive (SRA301985). Code is available from <https://github.com/justinjee/MDS> and https://github.com/susinmotion/barcode_tries.

31. Berend, D. & Tassa, T. Improved bounds on Bell numbers and on moments of sums of random variables. *Probab. Math. Stat.* **30**, 185–205 (2010).
32. Woods, R. J. *et al.* Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436 (2011).
33. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
34. Sarkar, S., Ma, W. T. & Sandri, G. H. On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**, 173–179 (1992).



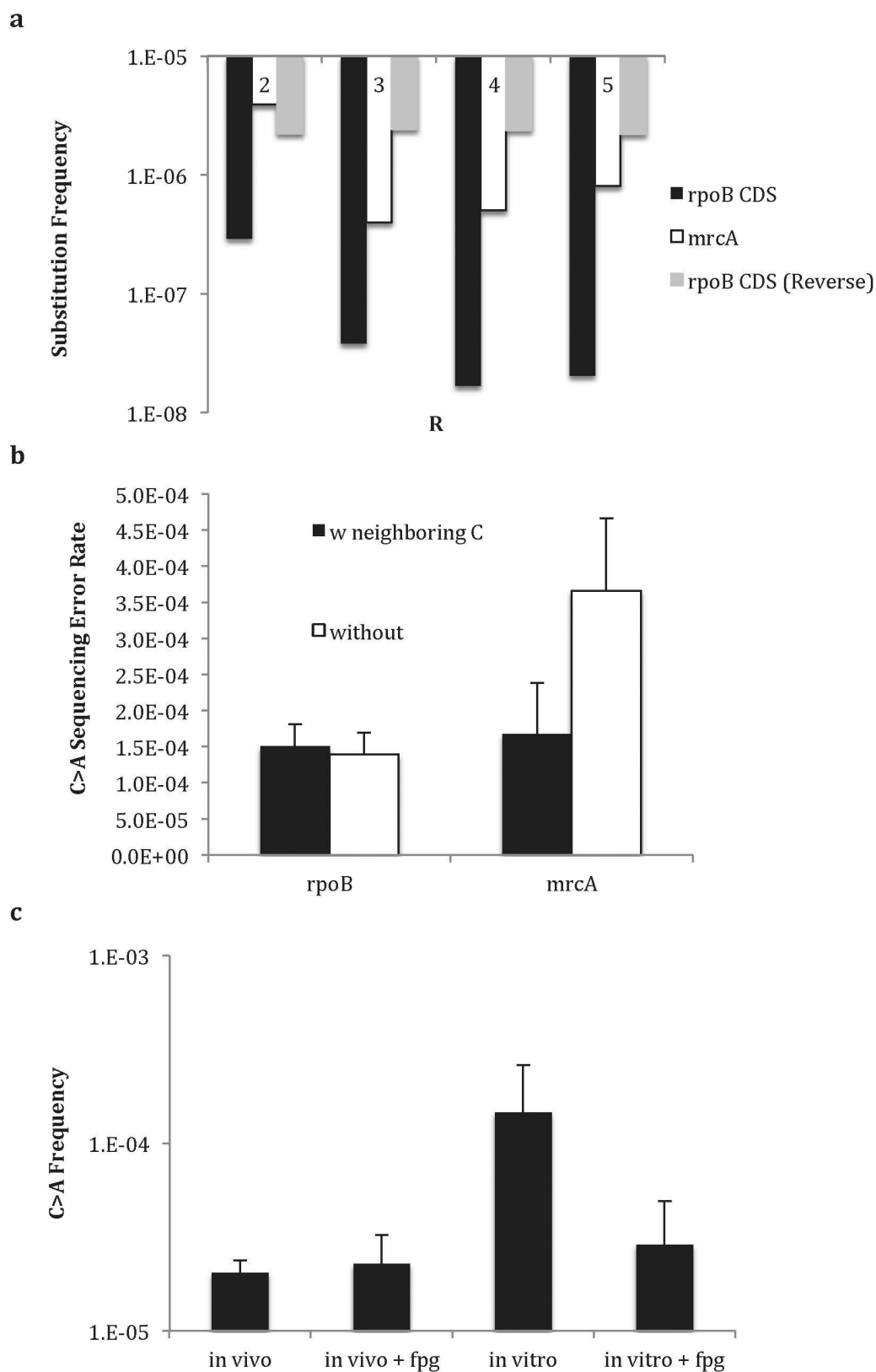
Extended Data Figure 1 | MDS accuracy and yield. a, Mock culture composed of *rpoB* point mutants of known concentration was sequenced using MDS. Output concentrations of each point mutant recovered from $R=2$ analysis are plotted against its input concentration (see Supplementary Information Table 2 for details). **b, c,** Distribution of the sizes of barcode families in four trials, shown as \log_{10} (number of barcode

families) per trial versus size of barcode family in reads (R). **b,** Trials used for the calibration run shown in **a** (~ 100 M reads total, divided into four trials). **c,** Representative quadruplicate trials (from *rpoB* of wild-type bacteria grown in LB broth with no antibiotics) taking up a total of one quarter of the output of a HiSeq rapid run, a total of ~ 60 M reads.



Extended Data Figure 2 | Dual-barcode MDS. **a**, Barcodes are attached to original DNA molecules as per MDS protocol. After linear amplification, a second barcode is attached to the opposite end of each read (see Supplementary Information: testing sample preparation and PCR efficiency). Exponential PCR is then performed. In the analysis phase, reads can be grouped both by primary barcode (that is, a classic MDS barcode family) and a second barcode corresponding to a 'subfamily' of reads with the same parent from a particular linear amplification step

before exponential amplification. **b**, The probability that for a given family only reads of one subfamily are recovered (a 'homogenous' barcode) decreases exponentially with R . For example, for $R = 3$, the probability all 3 reads are of the same subfamily is 0.02. **c**, We show the number of reads in each subfamily, sorted within each column by subfamily size, for the 1,500 largest primary barcode families in the experiment. For families of such size, it is unlikely that a single subfamily will account for more than 25% of the total number of reads recovered from that family.



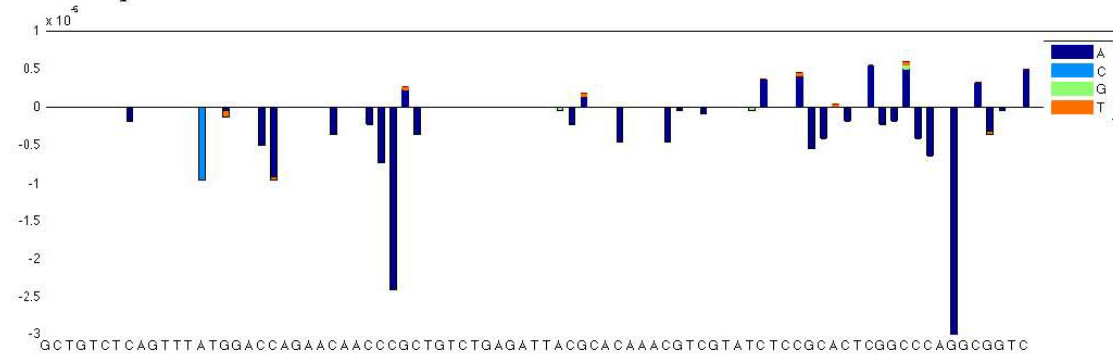
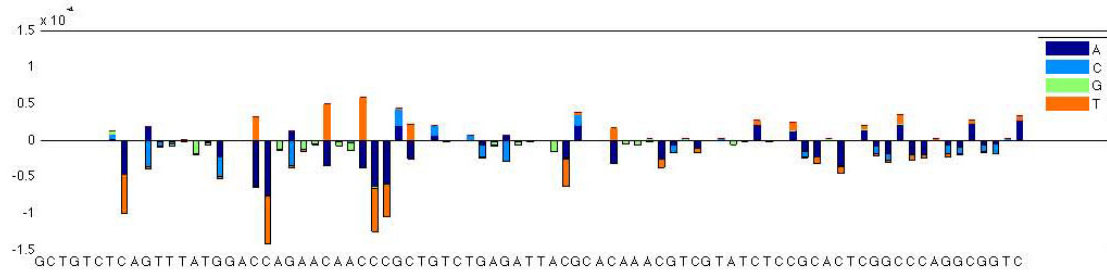
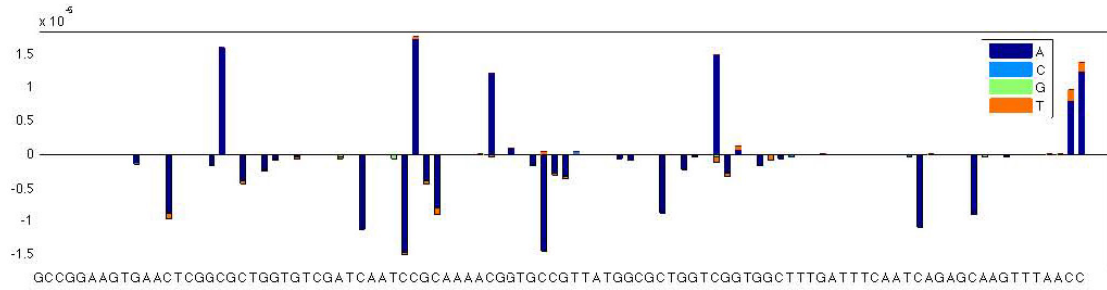
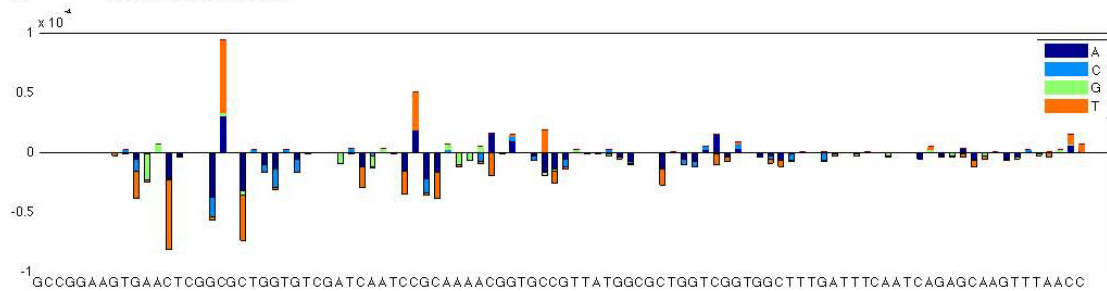
Extended Data Figure 3 | Substitution frequency controls.

a, Empirically, average substitution frequency (with high frequency substitutions such as C→A excluded) stabilizes as R increases. Note, substitution frequencies are not normalized by number of generations.

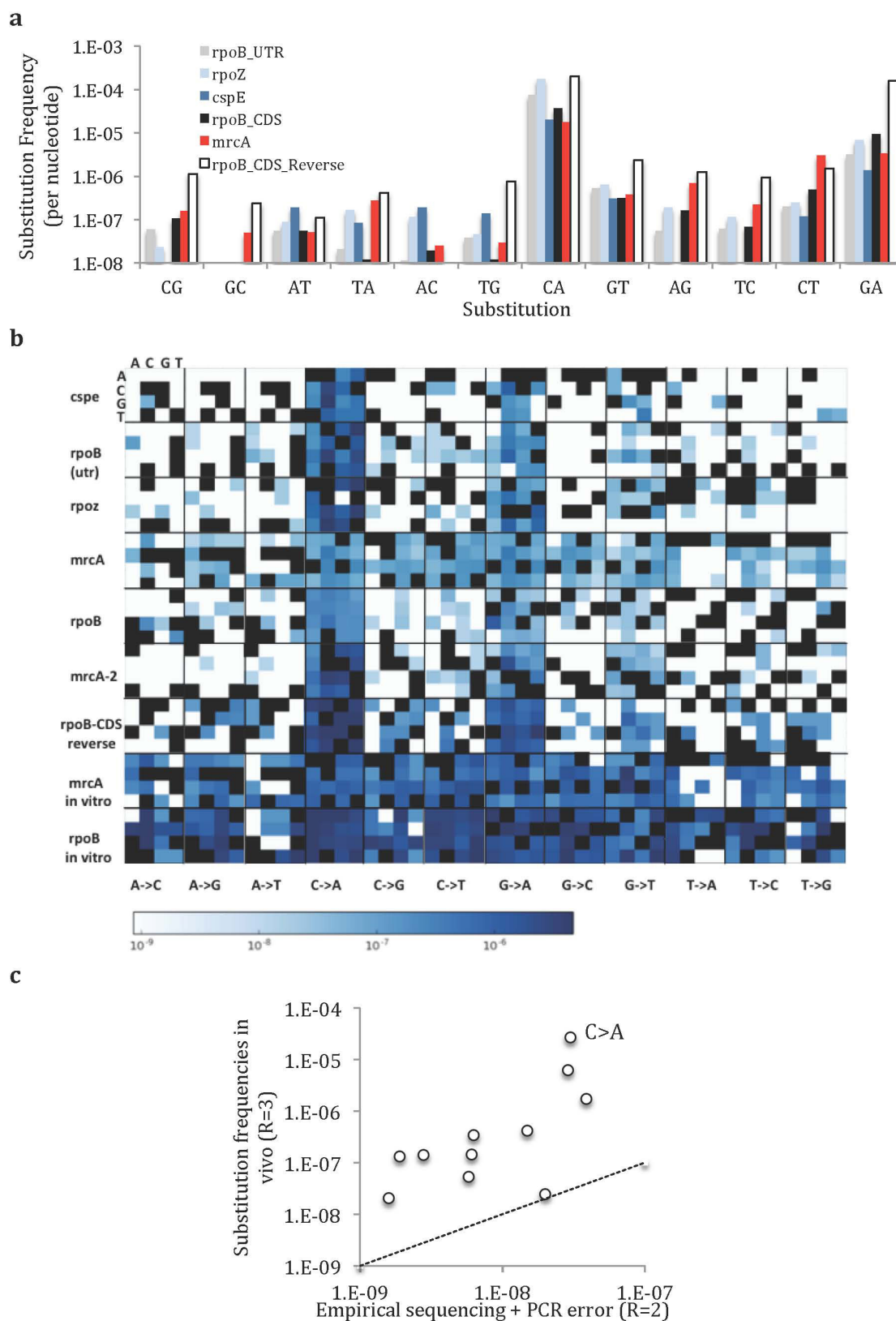
b, Empirical sequencing C→A error rate at C→A mutational hotspots with neighbouring Cs (same as those in Fig. 3c) versus all other positions.

c, C→A substitution frequencies when 10% 8-oxoG is synthetically added

to *in vitro* DNA and in FPG-treated samples. Frequencies are reported from ROI positions with potential 8-oxoG incorporations as described in template 'rpoB_reverse_complement_8-oxo-Dg'. Frequencies are reported at $R = 2$ level. For $R > 2$, no C→A substitutions were found in 72,646 *in vitro* template sites. Data represent biological triplicates. Error bars are standard deviation.

a *rpoB* CDS in vivo**b** *rpoB* CDS in vitro**c** *mrcA* in vivo**d** *mrcA* in vitro

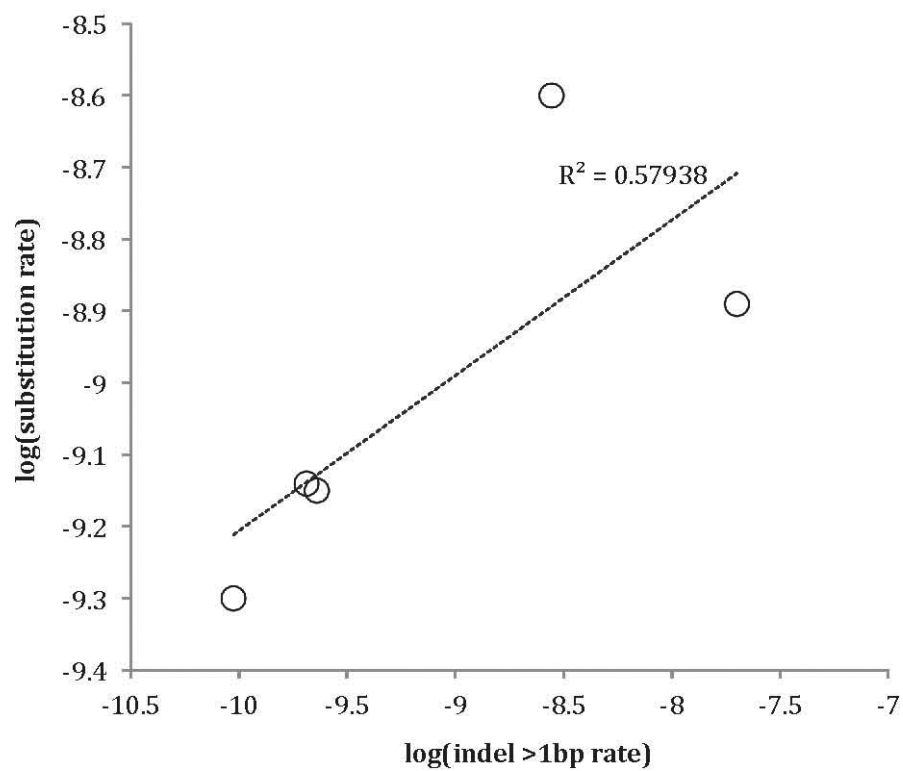
Extended Data Figure 4 | Substitution rates per locus. Positive frequencies denote synonymous substitutions. Negative frequencies denote nonsynonymous substitutions. **a, c,** Values are averaged across quadruplicate trials. **b, d,** *In vitro* synthesized DNA has undergone 20-cycle PCR amplification using Q5 polymerase.



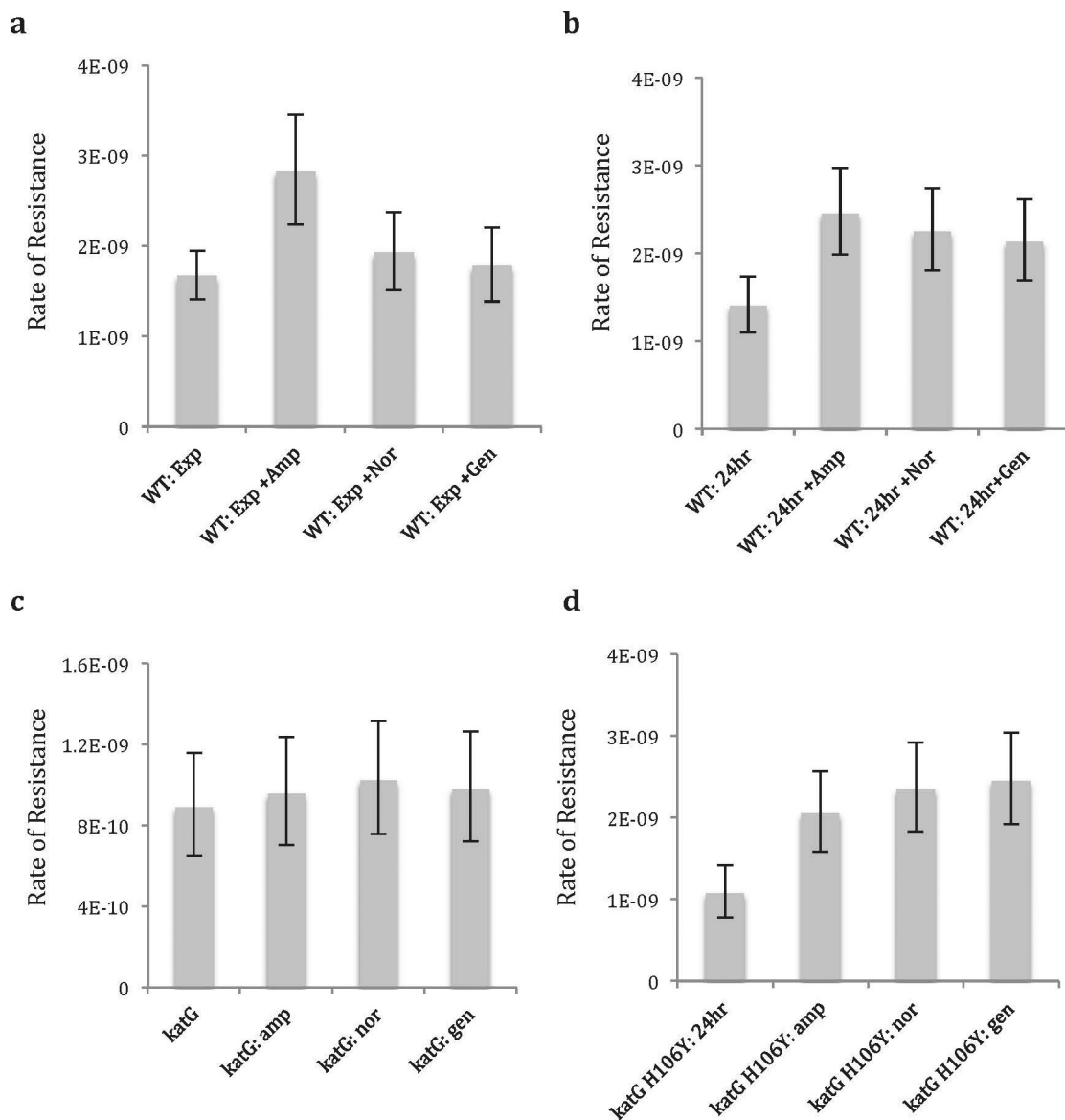
Extended Data Figure 5 | Mutational spectra and contexts.

a, Substitution frequencies of all ROIs after ~ 120 generations of growth. Note that values are not normalized for the number of generations and are thus true frequencies, rather than mutation rates. **b**, Mutation frequencies are shown in context of their 5' (A, C, G, or T on the x axis) and 3' (A, C, G, or T on the y axis) neighbours. **c**, The relative relationship between *in vivo* substitution frequencies and expected errors due to

sequencing and PCR (from *in vitro* DNA assays) is poorly described by a linear approximation ($R^2 = 0.27$). Furthermore, the recovered frequency from *in vivo* substitutions ($R = 3$) is higher than the rate of error (equivalent frequencies would be represented by the dotted line), even with the relatively relaxed read-cutoff threshold of $R = 2$ (the sequencing + PCR error with an $R = 3$ cutoff is approximately an order of magnitude lower). Templates are *rpoB* CDS and *mrcA* ROIs.

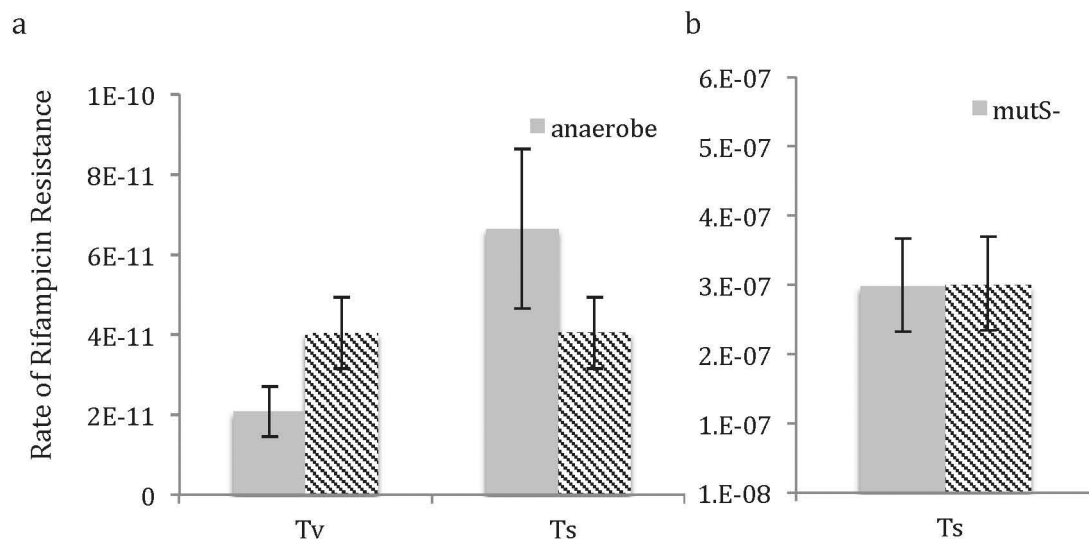


Extended Data Figure 6 | Comparing substitution rate and indel rate across 5 ROIs reveals a positive correlation. Pearson correlation coefficient = 0.76.

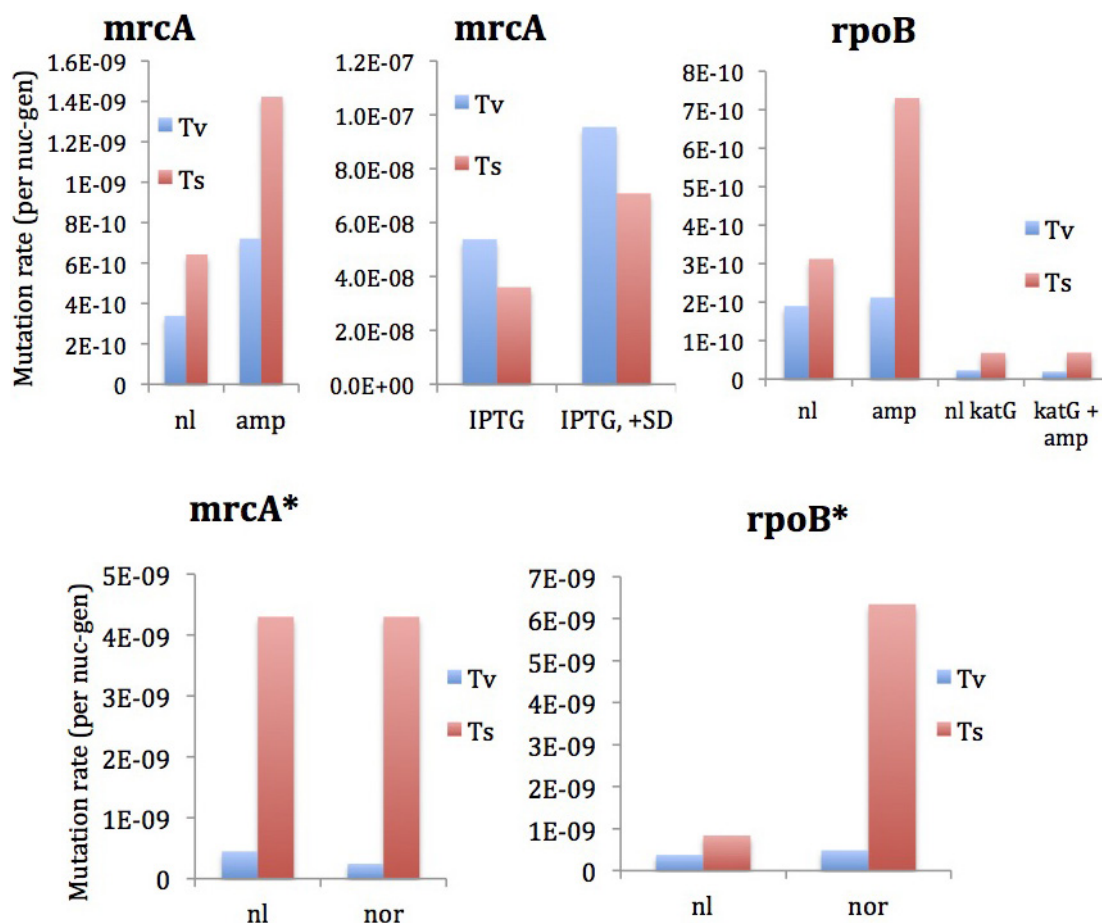


Extended Data Figure 7 | Rate of rifampicin resistance per generation. **a–d**, As calculated in fluctuation assays in wild-type cells grown in exponential phase only (**a**), wild-type cells grown to saturation (**b**), katG overexpression mutant grown to saturation (**c**) and inactive katG (H106Y

point mutation) overexpression mutant grown to saturation (**d**). Growth in LB broth was supplemented with possible subinhibitory doses of ampicillin (amp), norfloxacin (nor), or gentamycin (gen). Rates are mean. Error bars are 95% CI. $N = 25$ (see Methods: fluctuation assays).

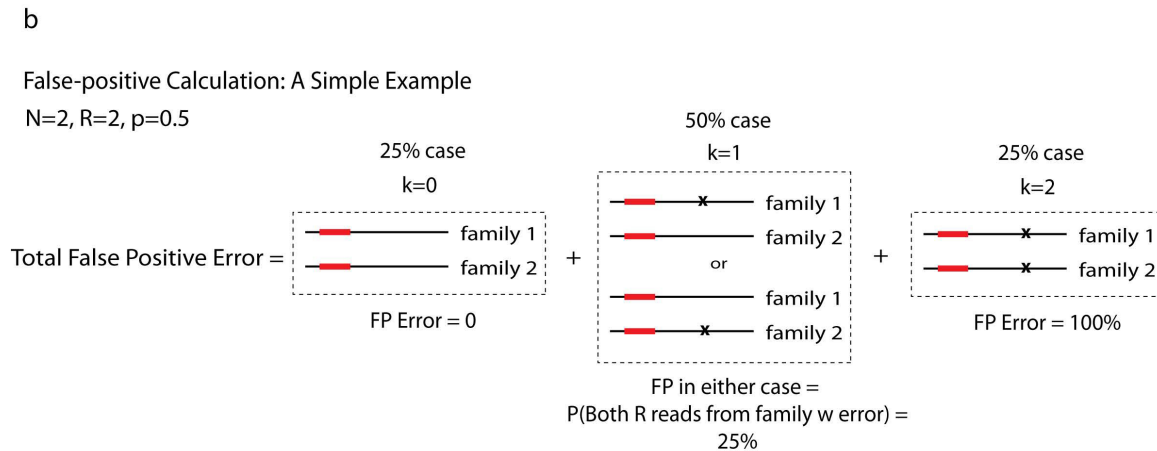
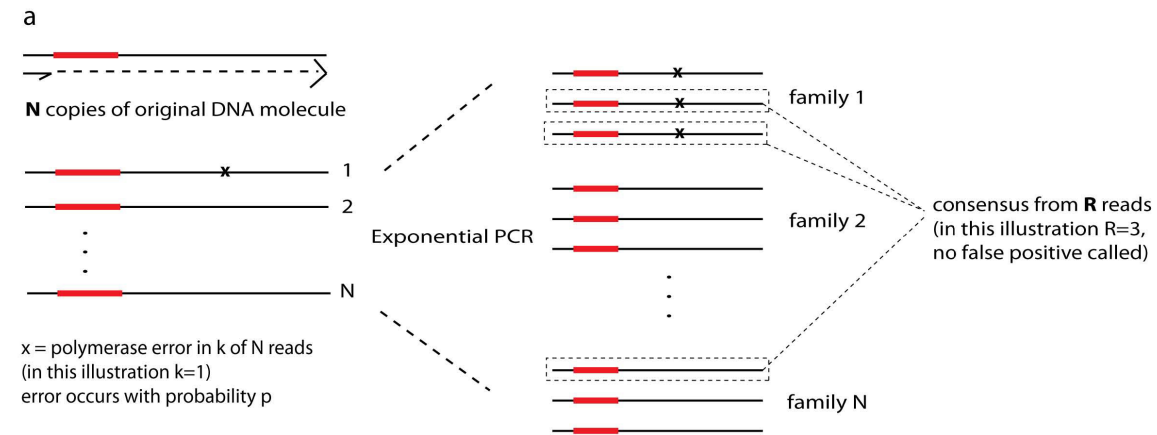


Extended Data Figure 8 | Transversion and transition rates (per nucleotide-generation). As calculated in fluctuation assays in anaerobic conditions (a) and in a *mutS* knockout (b). Note that because the transition (Ts) rate was high in *MutS* strains, transversion mutations could not be detected. Rates are mean. Error bars are 95% CI. $N = 25$ (see Methods: fluctuation assays).



Extended Data Figure 9 | Rates of *rpoB* and *mrcA* substitutions in the presence of antibiotics as calculated by MDS. Asterisks indicate cultures grown separately and prepared with Phusion rather than Q5. Although not shown, we note that only in-frame (3×) indels were observed in

rpoB in fluctuation assays, as expected since frameshift indels would be deleterious. These increased in frequency by a factor of 10 on addition of norfloxacin.



Extended Data Figure 10 | Schematic depicting the mathematical derivation of the false positive rate of MDS due to polymerase error.

a, The origin of various terms used in equations (2)–(7). **b**, Illustration of an example calculation of false positive rate given more ‘intuitive’ values of N , R and P . The false positive rate is calculated in a way that accounts for the possibility that an error in one or more ‘linear’ cycles propagates to

a whole family of reads. The number of reads with an error (k) is Poisson distributed according to equation (2). The probability of a false positive is the sum of the probabilities that all R reads come from one of k families, for all possible k , according to equation (3). Note that in practice, $P < 10^{-6}$, and in our study $N = 12$, $R > 2$, making the false positive rate much lower (see Fig. 1).

Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion

Franziska Faber¹, Lisa Tran¹, Mariana X. Byndloss¹, Christopher A. Lopez¹, Eric M. Velazquez¹, Tobias Kerrinnes¹, Sean-Paul Nuccio¹, Tamding Wangdi¹, Oliver Fiehn^{2,3}, Renée M. Tsoilis¹ & Andreas J. Bäuml¹

Changes in the gut microbiota may underpin many human diseases, but the mechanisms that are responsible for altering microbial communities remain poorly understood. Antibiotic usage elevates the risk of contracting gastroenteritis caused by *Salmonella enterica* serovars¹, increases the duration for which patients shed the pathogen in their faeces, and may on occasion produce a bacteriologic and symptomatic relapse^{2,3}. These antibiotic-induced changes in the gut microbiota can be studied in mice, in which the disruption of a balanced microbial community by treatment with the antibiotic streptomycin leads to an expansion of *S. enterica* serovars in the large bowel⁴. However, the mechanisms by which streptomycin treatment drives an expansion of *S. enterica* serovars are not fully resolved. Here we show that host-mediated oxidation of galactose and glucose promotes post-antibiotic expansion of *S. enterica* serovar Typhimurium (*S. Typhimurium*). By elevating expression of the gene encoding inducible nitric oxide synthase (iNOS) in the caecal mucosa, streptomycin treatment increased post-antibiotic availability of the oxidation products galactarate and glucarate in the murine caecum. *S. Typhimurium* used galactarate and glucarate within the gut lumen of streptomycin pre-treated mice, and genetic ablation of the respective catabolic pathways reduced *S. Typhimurium* competitiveness. Our results identify host-mediated oxidation of carbohydrates in the gut as a mechanism for post-antibiotic pathogen expansion.

A recent *in silico* analysis suggests that pathways involved in galactarate uptake and catabolism are associated with *S. enterica* serovars that cause gastrointestinal disease⁵. Galactarate fermentation is one of the biochemical reactions used to differentiate members of the genus *Salmonella* into serovars. Although 98.2% of serovars associated with gastrointestinal infections can ferment this carbon source, only 15.4% of serovars associated with extraintestinal disease test positive for this reaction⁶ (Extended Data Fig. 1a). However, the biological significance of this association is not clear, because galactarate is a xenobiotic that is not normally produced by mammals or expected to be present within the diet. We therefore investigated the origin of galactarate in the intestine.

Consistent with the idea that galactarate is a xenobiotic, the concentration of this sugar in mouse chow was very low, as suggested by gas chromatography/mass spectrometry (GC/MS) measurements (Extended Data Fig. 1b). To investigate whether this nutrient is normally available to promote growth in mucus, we constructed a *S. Typhimurium* strain lacking the *gudT ygcY gudD STM2959* operon (*gudT-STM2959* mutant, Extended Data Fig. 1c), which encodes proteins involved in galactarate uptake and catabolism⁷. Expression of the *gudT ygcY gudD STM2959* operon in *S. Typhimurium* is induced by hydrogen, a fermentation product of the gut microbiota⁸. Deletion of galactarate utilization genes rendered *S. Typhimurium* unable to ferment galactarate and glucarate, but did not affect its ability to utilize other monosaccharides (Fig. 1a). Genetic ablation of the

galactarate/glucarate utilization genes did not reduce the fitness of *S. Typhimurium* for anaerobic growth on hog mucin as the sole carbon source, but fitness of the *gudT-STM2959* mutant was reduced compared to the wild type when galactarate or glucarate was added to the medium (Fig. 1b). These data suggested that neither the diet nor the mucus naturally contained biologically relevant quantities of a substrate for enzymes encoded by the *gudT ygcY gudD STM2959* operon.

We next investigated the contribution of the *gudT ygcY gudD STM2959* operon to post-antibiotic pathogen expansion. Treatment of mice with a single dose of streptomycin one day before infection (pre-treatment with streptomycin) increased recovery of the wild-type *S. Typhimurium* from the colon contents of mice by approximately one order of magnitude compared to animals that had not received antibiotics ($P < 0.05$) (Fig. 1c). Genetic ablation of the galactarate/glucarate utilization genes significantly ($P < 0.05$) reduced recovery of *S. Typhimurium* from streptomycin pre-treated mice, but not from mice that had not received antibiotics. Genetic complementation with a plasmid carrying the cloned *gudT ygcY gudD STM2959* genes restored recovery of the *gudT-STM2959* mutant from streptomycin pre-treated mice to levels observed with the wild-type *S. Typhimurium*. Collectively, these data provided genetic evidence for a contribution of the *gudT ygcY gudD STM2959* operon to post-antibiotic expansion of *S. Typhimurium*.

Preconditioning of mice with streptomycin increases the severity of *S. Typhimurium* induced colitis⁹. We therefore investigated whether the availability of galactarate and/or glucarate is elevated during severe colitis, a host response triggered by the action of two type III secretion systems (T3SS-1 and T3SS-2), which constitute the main virulence factors of *S. Typhimurium*^{9,10}. To prevent the generation of acute intestinal inflammation, we used avirulent *S. Typhimurium* strains lacking a functional T3SS-1 (due to a mutation in *invA*) and T3SS-2 (due to a mutation in *spiB*). Streptomycin pre-treated mice were infected either with a 1:1 mixture of the wild-type bacteria and a *gudT-STM2959* mutant or with a 1:1 mixture of an *invA spiB* mutant and an *invA spiB gudT-STM2959* mutant. In each competition, the galactarate/glucarate utilization-proficient strain (the wild-type bacteria or the *invA spiB* mutant) was recovered in higher numbers than the corresponding galactarate/glucarate utilization-deficient strain (the *gudT-STM2959* mutant or the *invA spiB gudT-STM2959* mutant) (Fig. 2a). However, only mice infected with a mixture of the wild-type bacteria and a *gudT-STM2959* mutant developed acute intestinal inflammation (Extended Data Fig. 2 and Extended Data Tables 1 and 2). When the experiment was repeated with mice that had not received streptomycin, the presence of genes for galactarate/glucarate utilization no longer conferred a fitness advantage (Fig. 2a). To distinguish between glucarate and galactarate as possible carbon sources, we inactivated *gudD*, encoding glucarate dehydratase, or *garD*, encoding galactarate dehydratase (Extended Data Fig. 1c). During *in vitro* growth, genetic ablation of *gudD* only reduced *S. Typhimurium* fitness in medium containing glucarate, while deletion of the *garD* gene only reduced fitness in

¹Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, One Shields Avenue, Davis, California 95616, USA. ²Genome Center, University of California at Davis, One Shields Avenue, Davis, California 95616, USA. ³King Abdulaziz University, Biochemistry Department, Jeddah 21412, Saudi Arabia.

fitness advantage conferred by the galactarate/glucarate utilization genes after streptomycin treatment was abrogated in *Nos2*-deficient mice (Fig. 2a). These data suggested that the *Nos2* gene was necessary to generate a substrate for enzymes encoded by the *gudT ygcY gudD STM2959* operon.

To determine whether the streptomycin-induced increase in the caecal galactarate and glucarate concentrations (Fig. 2b, c) was *Nos2*-dependent, we measured galactarate and glucarate concentrations in caecal contents from *Nos2*-deficient mice four days after streptomycin treatment by GC/MS. Strikingly, streptomycin treatment did not increase the availability of these sugars in *Nos2*-deficient mice (Fig. 2b, c). These data further supported the idea that generation of post-antibiotic galactarate and glucarate required an intact *Nos2* gene.

To investigate whether the fitness advantage conferred by galactarate/glucarate utilization required iNOS activity, streptomycin-treated mice (C57BL/6) received drinking water supplemented with aminoguanidine hydrochloride, a specific iNOS inhibitor¹³, and were infected with a 1:1 mixture of the wild-type *S. Typhimurium* and a *gudT-STM2959* mutant. Inhibition of iNOS activity with AG significantly ($P < 0.05$) blunted the fitness advantage conferred by the galactarate/glucarate utilization genes (Fig. 2a), suggesting that iNOS activity was required for generating galactarate and glucarate in the large intestine.

The host enzyme iNOS uses L-arginine to produce nitric oxide (NO), a reactive nitrogen species¹⁴. Reactive nitrogen species are known catalysts in the oxidation of alcohols and aldehydes¹⁵. We thus hypothesized that by generating reactive nitrogen species, streptomycin-induced iNOS synthesis might drive an oxidation of monosaccharides, thereby yielding the oxidation products galactarate and glucarate. To investigate whether reactive nitrogen species might oxidize galactose and glucose to galactarate and glucarate, respectively, we used 2,2,6,6-tetramethyl piperidine-1-oxyl (TEMPO), which is a stable free nitroxyl radical that mimics the activity of reactive nitrogen species (reviewed in ref. 16). Galactose and glucose were incubated in the presence of TEMPO and a co-oxidant (NaOCl) or the co-oxidant alone. Detection by GC/MS indicated that TEMPO oxidized galactose to galactarate (Fig. 2d) and glucose to glucarate (Fig. 2e). Next, we investigated whether the monosaccharides galactose and glucose were present in caecal contents. Galactose and glucose were detected in caecal contents of conventional (C57BL/6) mice, *Nos2*-deficient mice and germ-free mice (Fig. 2f, g). Collectively, these data suggested that monosaccharides were present in the murine caecum and could be oxidized by reactive nitrogen species to yield galactarate and glucarate.

Gene clusters for the utilization of galactarate and glucarate are also present in *Escherichia coli* and other related Enterobacteriaceae (Extended Data Fig. 1c). As treatment with streptomycin leads to an uncontrolled expansion of *E. coli* in the murine intestine¹⁷, we investigated whether the underlying mechanism also involved utilization of galactarate and glucarate. To test this, we deleted the *gudDXP* and *garD* genes in the human *E. coli* isolate Nissle 1917 (Extended Data Fig. 1c). Deletion of the *gudDXP garD* genes rendered *E. coli* unable to grow with galactarate or glucarate as the sole carbon source, but did not affect its ability to utilize glycerate (Extended Data Fig. 5a). The *gudDXP garD* genes conferred a fitness advantage during growth of *E. coli* in the colon of streptomycin pre-treated mice, which was significantly ($P < 0.05$) diminished after treatment with the iNOS inhibitor aminoguanidine hydrochloride (Extended Data Fig. 5b).

Here we show that by inducing the production of host-derived reactive nitrogen species, streptomycin treatment generates galactarate and glucarate in the gut lumen, thereby providing *S. Typhimurium* and *E. coli* with a considerable fitness advantage. Increases in galactarate and glucarate levels are also observed after treatment of mice with cefoperazone or a cocktail of vancomycin and bacitracin^{18,19}. A post-antibiotic expansion of Enterobacteriaceae is of concern due to the recent emergence of carbapenem antibiotic resistance within this group. Exposure of patients in intensive care units to broad-spectrum antibiotics is a known risk factor for acquiring an infection with

carbapenem-resistant *E. coli* and *Klebsiella* isolates²⁰. Our findings identify host-mediated sugar oxidation as a new mechanism contributing to post-antibiotic expansion of Enterobacteriaceae.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 September 2015; accepted 24 May 2016.

Published online 15 June 2016.

- Pavia, A. T. *et al.* Epidemiologic evidence that prior antimicrobial exposure decreases resistance to infection by antimicrobial-sensitive *Salmonella*. *J. Infect. Dis.* **161**, 255–260 (1990).
- Nelson, J. D., Kusmiesz, H., Jackson, L. H. & Woodman, E. Treatment of *Salmonella* gastroenteritis with ampicillin, amoxicillin, or placebo. *Pediatrics* **65**, 1125–1130 (1980).
- Aserkoff, B. & Bennett, J. V. Effect of antibiotic therapy in acute salmonellosis on the fecal excretion of salmonellae. *N. Engl. J. Med.* **281**, 636–640 (1969).
- Bohnhoff, M., Drake, B. L. & Miller, C. P. Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* infection. *Proc. Soc. Exp. Biol. Med.* **86**, 132–137 (1954).
- Nuccio, S. P. & Bäuml, A. J. Comparative analysis of *Salmonella* genomes identifies a metabolic network for escalating growth in the inflamed gut. *MBio* **5**, e00929–14 (2014).
- Kelterborn, E. *Kauffmann–White–Schema* (1989) 1–171 (Bundesgesundheitsamt, 1992).
- Lamichhane-Khadka, R., Frye, J. G., Porwollik, S., McClelland, M. & Maier, R. J. Hydrogen-stimulated carbon acquisition and conservation in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **193**, 5824–5832 (2011).
- Lamichhane-Khadka, R., Benoit, S. L., Maier, S. E. & Maier, R. J. A link between gut community metabolism and pathogenesis: molecular hydrogen-stimulated glucarate catabolism aids *Salmonella* virulence. *Open Biol.* **3**, 130146 (2013).
- Barthel, M. *et al.* Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect. Immun.* **71**, 2839–2858 (2003).
- Isolis, R. M., Adams, L. G., Ficht, T. A. & Bäuml, A. J. Contribution of *Salmonella typhimurium* virulence factors to diarrheal disease in calves. *Infect. Immun.* **67**, 4879–4885 (1999).
- Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013).
- Spees, A. M. *et al.* Streptomycin-induced inflammation enhances *Escherichia coli* gut colonization through nitrate respiration. *MBio* **4**, e00430–13 (2013).
- Stefanovic-Racic, M. *et al.* Comparison of the nitric oxide synthase inhibitors methylarginine and aminoguanidine as prophylactic and therapeutic agents in rat adjuvant arthritis. *J. Rheumatol.* **22**, 1922–1928 (1995).
- Palmer, R. M., Ashton, D. S. & Moncada, S. Vascular endothelial cells synthesize nitric oxide from L-arginine. *Nature* **333**, 664–666 (1988).
- de Souza, M. The use of TEMPO (2,2,6,6-tetramethylpiperidine-N-oxyl) for the oxidation of primary and secondary alcohols. *Quim. Nova* **27**, 287–292 (2004).
- Sheldon, R. A. & Arends, I. W. C. E. Organocatalytic oxidations mediated by nitroxyl radicals. *Adv. Synth. Catal.* **346**, 1051–1071 (2004).
- Ozawa, A. & Freter, R. Ecological mechanism controlling growth of *Escherichia coli* in continuous flow cultures and in the mouse intestine. *J. Infect. Dis.* **114**, 235–242 (1964).
- Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nature Commun.* **5**, 3114 (2014).
- Hwang, I. *et al.* Alteration of gut microbiota by vancomycin and bacitracin improves insulin resistance via glucagon-like peptide 1 in diet-induced obesity. *FASEB J.* **29**, 2397–2411 (2015).
- Tängdén, T. & Giske, C. G. Global dissemination of extensively drug-resistant carbapenemase-producing Enterobacteriaceae: clinical perspectives on detection, treatment and infection control. *J. Intern. Med.* **277**, 501–512 (2015).

Acknowledgements This work was supported by Public Health Service grants OD010931 (E.M.V.), AI060555 (S.-P.N.), AI096528 (A.J.B.), AI109799 (R.M.T.), AI112241 (C.A.L.), AI112258 (R.M.T.), AI112445 (A.J.B.), U24 DK097154 (O.F.), AI112949 (A.J.B.) and AI114922 (A.J.B.).

Author Contributions F.F. performed bacterial growth assays, most animal experiments and analysed the results. O.F. performed GC/MS measurements. M.X.B. scored histological sections. A.J.B., L.T., C.A.L., E.M.V., T.K. and T.W. assisted with animal experiments. F.F., S.-P.N., R.M.T. and A.J.B. were responsible for the overall study design. F.F. and A.J.B. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J.B. (ajbaumler@ucdavis.edu).

Reviewer Information *Nature* thanks D. Bolam, D. Bumann, M. Fischbach and D. M. Monack for their contribution to the peer review of this work.

METHODS

Bacterial strains and growth conditions. *S. Typhimurium* and *E. coli* strains used in this study are listed in Extended Data Table 3. All cultures were routinely grown aerobically at 37 °C in either Luria-Bertani (LB) broth (10 g per litre tryptone, 5 g per litre yeast extract, 10 g per litre NaCl) or on LB agar plates (1.5% Difco agar) unless indicated otherwise. When necessary, antibiotics were added to the medium at the following concentrations: nalidixic acid (Nal) 50 mg per litre, kanamycin (Km) 100 mg per litre, chloramphenicol (Cm) 30 mg per litre, carbenicillin (Carb) 100 mg per litre.

Sugar fermentation assay. 5 ml of fermentation broth (peptone, 10 g per litre; bromothymol blue, 0.024 g per litre; final pH 7.4 ± 0.1) supplemented with the indicated carbon source (galactarate, glucarate, glucose, galactose, mannose or rhamnose, 10 g per litre each) or the control broth (no sugar added) were inoculated with 10 µl of an overnight culture of each indicated *S. Typhimurium* strain and incubated statically at 37 °C for 24 h. Fermentation of the sugar in the broth is indicated by a colour change from blue to yellow.

Anaerobic growth assays. 10 ml of M9 minimal medium (75 g per litre $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 30 g per litre KH_2PO_4 , 5 g per litre NaCl, 10 g per litre NH_4Cl , 0.1 mM CaCl_2 , 1 mM MgSO_4 , 0.001% thiamine) supplemented with hog mucin (0.1% w/v) or galactarate (0.04% w/v when added as sole carbon source, 0.1% w/v and 0.01% w/v when added to mucin broth) were inoculated with 2×10^3 colony-forming units (CFU) of each strain and incubated anaerobically at 37 °C for 24 h inside an anaerobic chamber (Bactron I Anaerobic Chamber; Sheldon Manufacturing, Cornelius). Bacterial numbers were determined by plating serial tenfold dilutions onto LB agar containing the appropriate antibiotics. The ratios of recovered wild-type and mutant bacteria after 24 h were normalized to the ratio at 0 h to calculate the competitive index.

Construction of plasmids. Standard cloning techniques were used to generate the plasmids used in this study. All plasmids and primers used in this study are listed in Extended Data Tables 4 and 5. PCR products were confirmed by sequencing (SeqWright Fisher Scientific, Houston). Suicide plasmids were propagated in *E. coli* DH5a λ pir. Plasmid pFF35 was constructed by PCR amplifying a 5' flanking fragment of *gudT* using primers 71 and 72 and a 3' flanking fragment *STM2959* using primers 73 and 74. The two PCR fragments were gel purified, digested with XbaI and ligated with T4 DNA ligase (NEB). The ligation mix served as a template for a PCR with primers 71 and 74 and the product was gel purified and cloned into pCR2.1 using the TOPO TA cloning kit (Invitrogen). The plasmid construct was confirmed by sequencing and designated pFF35. To generate a suicide plasmid for replacing the *gudT ygcY gudD STM2959* genes with a kanamycin (KSAC) cassette, the insert of plasmid pFF35 was excised using BamHI and ligated into the BamHI site of the suicide plasmid pRDH10. The resulting plasmid was digested with XbaI and ligated with a KSAC cassette generated from pBS34 by digestion with XbaI. The resulting suicide plasmid was designated pFF57.

To construct plasmids pFF62 and pFF63, respectively, chromosomal regions upstream and downstream of *gudD* and *garD* in IR715 were amplified by PCR and cloned into BamHI-digested pRDH10 using Gibson Assembly Master Mix (NEB). To construct plasmids pFF64 and pFF65, respectively, chromosomal regions upstream and downstream of the *gudDXP* operon and of *garD* from *E. coli* Nissle 1917 were amplified by PCR and cloned into BamHI digested pRDH10 using Gibson Assembly Master Mix (NEB).

For complementation of the *gudT-STM2959* mutant, the *gudT ygcY gudD STM2959* operon including its promoter region was PCR amplified using primers 92 and 93 or 94 and 95. The two PCR fragments were gel purified and cloned into BamHI digested pWSK29 using Gibson Assembly Master Mix (NEB). The complementation plasmid was verified by sequencing and designated pGUDT.

Construction of mutants in *S. Typhimurium*. All suicide plasmids were introduced into *S. Typhimurium* IR715 recipient strains by conjugation using *E. coli* S17-1 λ pir as the donor strain. Exconjugants were selected on LB + Nal + Cm to select for clones that had integrated the suicide plasmid. Sucrose counter-selection was performed as published previously²¹. Strains that were sucrose resistant and Cm^{S} were verified by PCR.

Plasmid pFF57 was introduced into FF7 and SPN487 to generate FF162 ($\Delta\text{gudT-STM2959::Km}^{\text{R}}$) and FF217 ($\Delta\text{invA } \Delta\text{spiB } \Delta\text{gudT-STM2959::Km}^{\text{R}}$), respectively. Plasmid pFF62 was introduced into AJB715 to generate FF464 (*phoN::Km^R invA ΔspiB gudD*). Plasmid pFF63 was introduced into AJB715 to generate strain FF461 (*phoN::Km^R ΔinvA ΔspiB garD*).

Construction of mutants in *E. coli* Nissle 1917. Suicide plasmids were introduced into *E. coli* Nissle 1917(pSW172) recipient strains by conjugation using *E. coli* S17-1 λ pir as the donor strain. To ensure stable propagation of pSW172, all steps of the conjugation were performed at 30 °C. Exconjugants were selected on LB + Carb + Cm to select for clones that had integrated the suicide plasmid. Subsequent sucrose counter-selection was performed as published previously²¹. Strains that were sucrose resistant and Cm^{S} were verified by PCR. If appropriate,

pSW172 was cured by growing the resulting mutant strains at 37 °C. Plasmids pFF64 and pFF65 were successively introduced into *E. coli* Nissle 1917(pSW172) to generate FF441 (*E. coli* Nissle 1917 *gudDXP garD*).

Generalized phage transduction. Phage P22 HT105/1 *int-201* was used for generalized transduction. A phage lysate of strain CS019 was used to transduce IR715 wild-type and the $\Delta\text{invA } \Delta\text{spiB}$ mutant (SPN487) to Cm^{R} generating FF176 (IR715 *phoN::Tn10d-Cam*) and FF183 (IR715 *phoN::Tn10d-Cam ΔinvA ΔspiB*), respectively. Transductants were cleaned from phage contaminations on Evans blue-Uranine (EBU) agar plates and tested for phage sensitivity by cross-streaking against P22 H5. The strains were verified by PCR to be *phoN::Tn10d-Cam*.

Animal experiments. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were blinded to allocation of mice for assessment of histopathology and readouts of inflammation. All animal experiments were approved by the Institutional Animal Care and Use Committees at the University of California, Davis. Female C57BL/6J wild-type and *Nos2*-deficient mice (B6.129P2-*Nos2^{tm1Lau/J}*) aged 9–12 weeks were obtained from The Jackson Laboratory (Bar Harbour). Mice were pre-treated with an oral dose of 20 mg streptomycin or sterile water 24 h before infection with *S. Typhimurium* or *E. coli* Nissle 1917, respectively. For single infections with *Salmonella*, mice were inoculated intra-gastrically with 2×10^8 CFU of the indicated *S. Typhimurium* strains. For competitive infections, mice were inoculated with a 1:1 mixture of each indicated strain. For infection with *E. coli* Nissle 1917, mice were infected with 2×10^9 CFU of a 1:1 mixture of the indicated strains. In some experiments, drinking water was supplemented with aminoguanidine hydrochloride (1 mg ml⁻¹). Mice were euthanized 4 days after infection. The colon contents were collected for enumeration of bacterial numbers, the distal caecum was collected for histopathology scoring. Bacterial numbers were determined by plating serial tenfold dilutions onto LB agar containing the appropriate antibiotics. The ratio of recovered wild-type and mutant bacteria in colon contents were normalized by the ratio in the inoculum to calculate the competitive index. For measurements of sugar concentrations in caecal contents, mice received an oral dose of either 20 mg streptomycin or sterile water. Mice were euthanized 4 days after streptomycin treatment and caecal contents were snap frozen in liquid nitrogen and stored at -80 °C. The distal caecal tissue was collected for RNA isolation.

Germ-free Swiss Webster mice were obtained from Taconic Farms. The mice were bred and housed under germ-free conditions inside gnotobiotic isolators (Park Bioservices, LLC). Weekly 16 S PCR and cultures were performed to evaluate the germ-free status of the mice. For experiments, male and female 6–8-week-old mice were transferred to a biosafety cabinet and maintained in sterile cages for the duration of the experiment. Caecal contents were snap frozen in liquid nitrogen and stored at -80 °C until further processing for metabolite measurements.

Quantitative real-time RT-PCR analysis. Murine caecal tissue was collected, snap frozen in liquid nitrogen and stored at -80 °C. Expression analysis was performed as described previously²². Briefly, RNA was extracted using TRI reagent (Molecular Research Center, Cincinnati) according to the manufacturer's protocol. Isolated RNA was DNase-treated (Applied Biosystems) and cDNA was synthesized from 1 µg of RNA using TaqMan reverse transcription reagents (Applied Biosystems). Real-time PCR was performed using SYBR-Green (Applied Biosystems) and the primers listed in Extended Data Table 5. The changes in mRNA levels of target genes were calculated using the comparative Ct method (Applied Biosystems) and normalized to the levels of *B2m* mRNA.

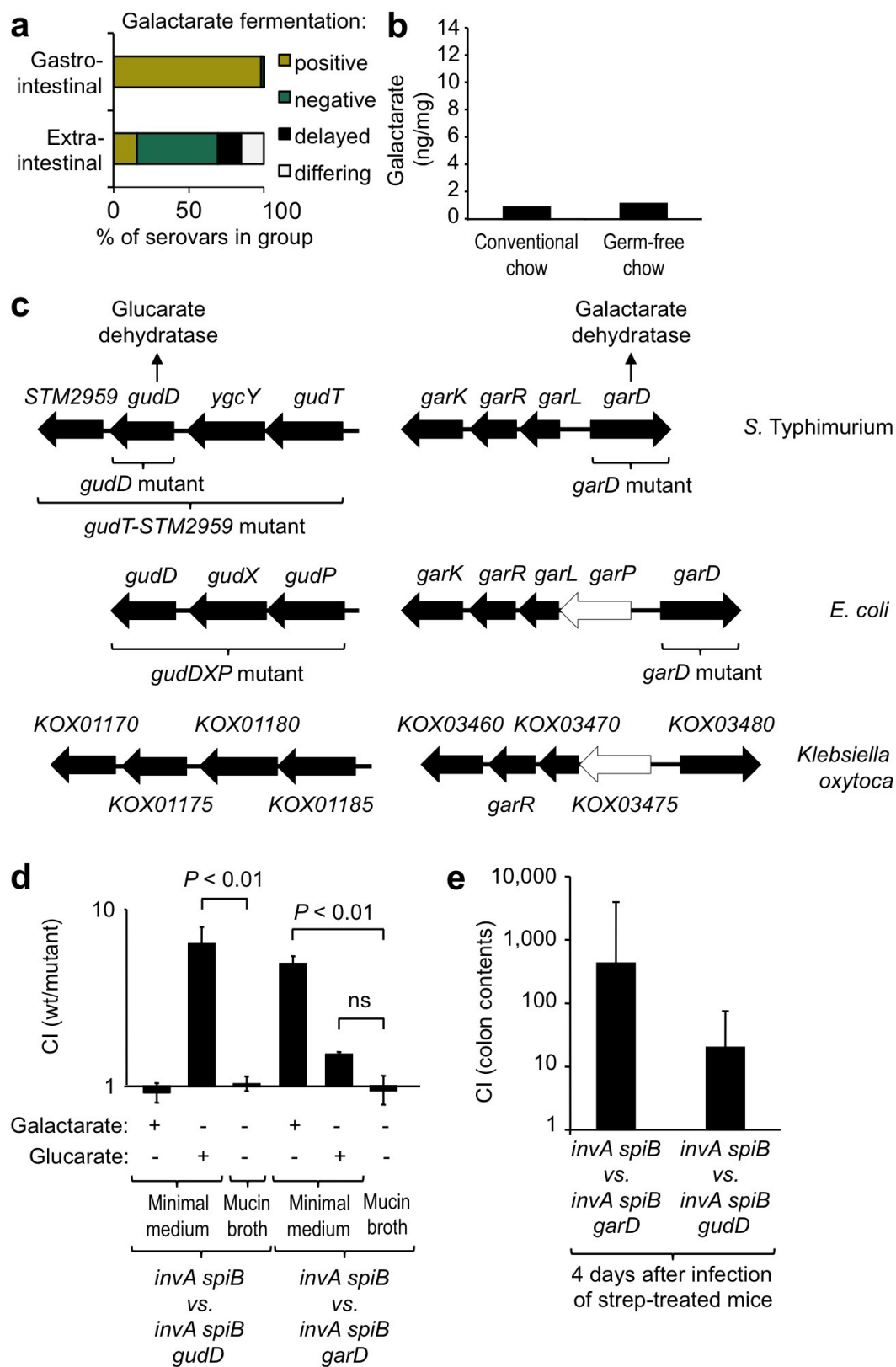
Histopathology. The distal caecal tissue was fixed in phosphate-buffered formalin and 5 µm sections of the tissue were stained with haematoxylin and eosin. Blinded scoring of tissue sections was performed by a veterinary pathologist based on the criteria listed in Extended Data Tables 1 and 2. Images were taken with a Zeiss Primo Star microscope with a 10× objective.

TEMPO-mediated oxidation of sugars. The oxidation reactions were carried out at room temperature in an open beaker fitted with a pH-meter electrode and a thermometer to monitor the reactions. The pH was kept constant at 8 by titration with a 0.5 M NaOH solution. D-galactose (2 g) and TEMPO (16 mg, 2,2,6,6-tetramethyl-1-piperidinyloxy, free radical, Sigma-Aldrich) were dissolved in 250 ml ultrapure water. As a control some of the oxidations were done without adding TEMPO to the reaction mixture. To start the oxidation, sodium hypochlorite solution (a total of 11 ml of 10–15% NaClO, Sigma-Aldrich) was added at a rate not exceeding a pH of 8.0. After adding the entire NaClO solution the pH was kept at 8.0 until no further pH shift was detectable. The reaction mixture was precipitated with 3 volumes of ≥ 95% ethanol at 4 °C for 48 h, filtered (Whatman qualitative filter paper, grade 602 h), washed with acetone and dried at 50 °C for 10–15 h. The resulting powdered mixture was used as a carbon source for *in vitro* growth assays and analysed by GC/MS for the presence of galactarate.

Measurement of sugar concentrations by GC/MS. Measurements were done at the West Coast Metabolomics Center at UC Davis as previously described²³. 20 mg of each sample, with D-Glucose-C-d7 added as the internal standard, were extracted with 1 ml of a pre-chilled acetonitrile:isopropanol:water (3:3:2) mixture. 450 µl aliquots of the supernatants were evaporated to dryness and subjected to a two-step derivatization using methoximation and trimethylsilylation. GC/MS analysis was performed using an Agilent 7890 Gas Chromatography system coupled to an Agilent 5977A Mass spectrometer. An Rtx-5Sil MS w/Integra-Guard column (30 m × 250 µm i.d., Restek), chemically bonded with a 1,4-bis(dimethyl-siloxy)phenylene-dimethyl polysiloxane cross-linked stationary phase (0.25 µm film thickness) was used to separate the derivatives. Helium was used as a carrier gas at a constant flow rate of 1.2 ml min⁻¹. The GC oven temperature was programmed to increase from 50 °C to 325 °C at a rate of 10 °C min⁻¹. The temperatures of the injector, transfer line, electron impact (EI) ion source, and quadrupole were set to 250 °C, 290 °C, 230 °C and 150 °C, respectively. The mass spectrometer was set to scan at a sampling rate of 4 and data was collected in a full scan mode (*m/z* 50 to 600). For quantification of sugars in the samples, a 6 point calibration curve was prepared with D-Glucose-C-d7 as internal standard. Agilent Mass Hunter quant software was used for data analysis.

Statistical analysis. The fold-changes of ratios for bacterial numbers and mRNA levels, respectively, and values for sugar concentrations were logarithmically transformed for statistical analysis. An unpaired Student's *t*-test was used to determine whether differences between groups were statistically significant (*P* < 0.05). Error bars indicate standard error of the mean (s.e.m).

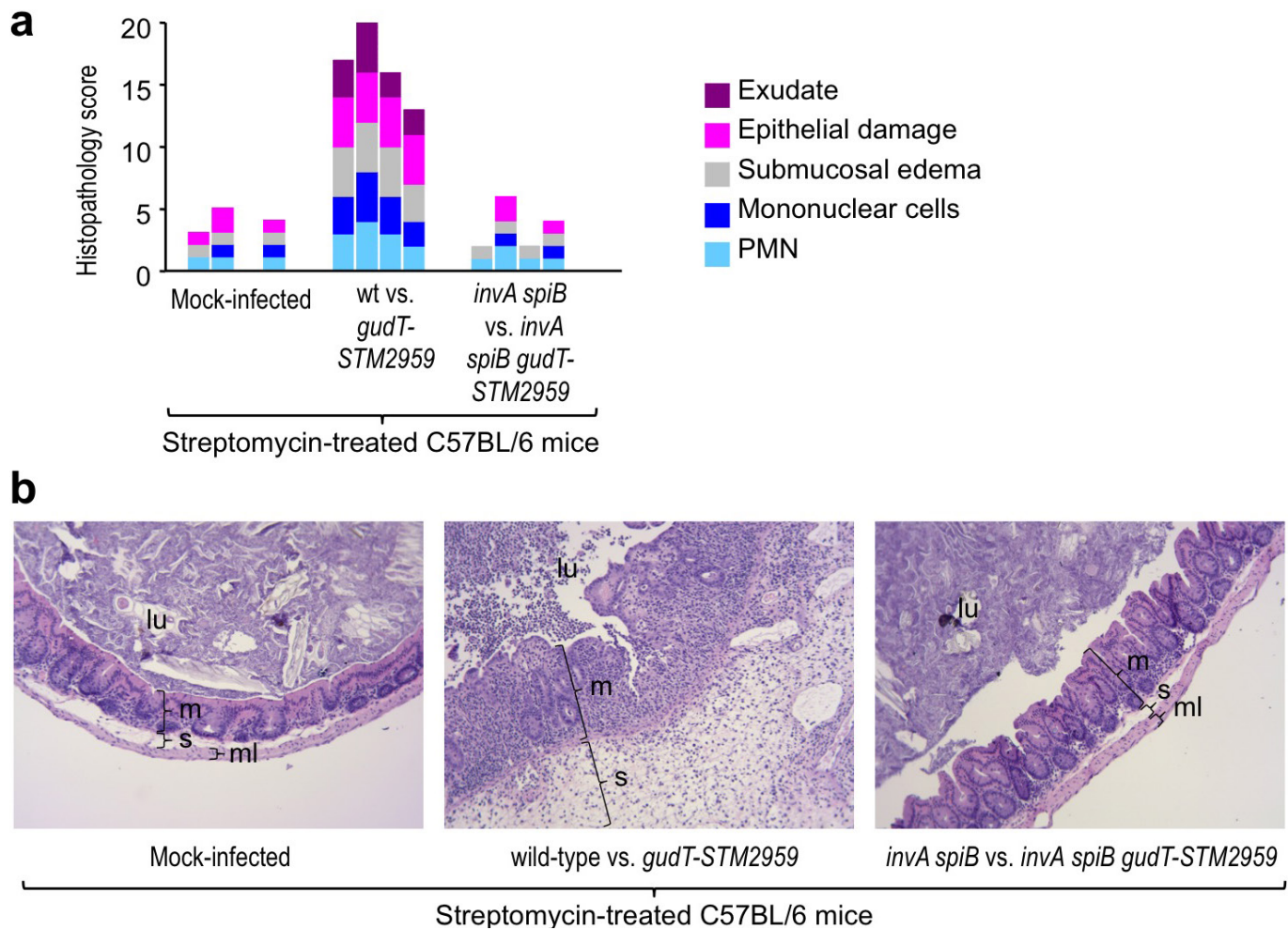
21. Lawes, M. & Maloy, S. MudSacl, a transposon with strong selectable and counterselectable markers: use for rapid mapping of chromosomal mutations in *Salmonella typhimurium*. *J. Bacteriol.* **177**, 1383–1387 (1995).
22. Winter, S. E. *et al.* Contribution of flagellin pattern recognition to intestinal inflammation during *Salmonella enterica* serotype Typhimurium infection. *Infect. Immun.* **77**, 1904–1916 (2009).
23. Fiehn, O. *et al.* Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J.* **53**, 691–704 (2008).
24. Grozdanov, L. *et al.* Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J. Bacteriol.* **186**, 5432–5441 (2004).
25. Stojiljkovic, I., Bäumler, A. J. & Heffron, F. Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the *cchA cchB eutE eutJ eutG eutH* gene cluster. *J. Bacteriol.* **177**, 1357–1366 (1995).
26. Miller, S. I., Kukral, A. M. & Mekalanos, J. J. A two-component regulatory system (*phoP phoQ*) controls *Salmonella typhimurium* virulence. *Proc. Natl Acad. Sci. USA* **86**, 5054–5058 (1989).
27. Rivera-Chávez, F. *et al.* *Salmonella* uses energy taxis to benefit from intestinal inflammation. *PLoS Pathog.* **9**, e1003267 (2013).
28. Kingsley, R. A. *et al.* Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica* serotype Typhimurium: identification of intestinal colonization and persistence determinants. *Infect. Immun.* **71**, 629–640 (2003).
29. Raffatellu, M. *et al.* Lipocalin-2 resistance confers an advantage to *Salmonella enterica* serotype Typhimurium for growth and survival in the inflamed intestine. *Cell Host Microbe* **5**, 476–486 (2009).
30. Winter, S. E. *et al.* Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. *Science* **339**, 708–711 (2013).



Extended Data Figure 1 | See next page for caption.

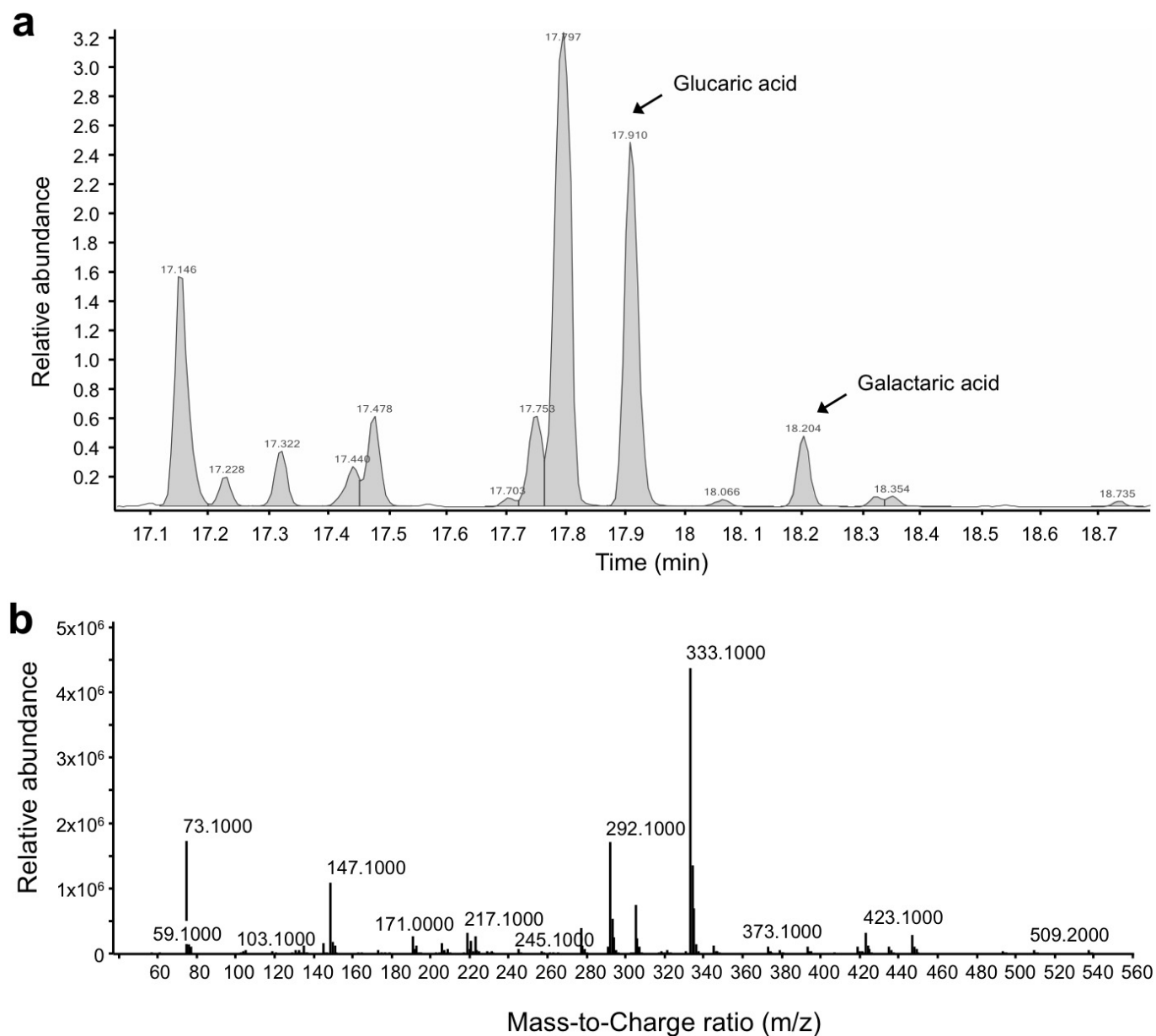
Extended Data Figure 1 | Galactarate/glucarate fermentation by *S. enterica*. **a**, One of the biochemical reactions used in the *Salmonella* serotyping scheme by Kauffman and White is the ability to ferment galactarate⁶. We divided 1,367 *S. enterica* subspecies *enterica* serovars into two groups: those associated with extraintestinal disease (serovars Typhi, Paratyphi A, Paratyphi B, Paratyphi C, Sendai, Choleraesuis, Typhisuis, Dublin, Bovismorbificans, Abortusovis, Abortusequi, Gallinarum biovar Gallinarum and Gallinarum biovar Pullorum) and those associated with human gastroenteritis (the remaining 1,354 serovars). The bar graph shows the percentages of serovars in each group that are positive, negative, delayed or differing (some isolates within the serovar are positive while others are negative) for this reaction. **b**, Detection of galactarate in chow for conventional or germ-free mice using GC/MS ($n = 4$).

c, Schematic drawing of the two gene clusters encoding proteins involved in the degradation of glucarate and galactarate in *S. Typhimurium* (ATCC14028), *E. coli* (Nissle 1917) and *Klebsiella oxytoca* (KCTC1686). Arrows indicate genes. The bracket indicates the DNA region deleted in the indicated mutants. **d**, Minimal medium or mucin broth supplemented with the indicated carbon sources (0.1% w/v) was inoculated with a 1:1 mixture of the wild-type *S. Typhimurium* and indicated mutants. Competitive index (CI) recovered after 24 h incubation in an anaerobic chamber. **e**, Streptomycin-treated C57BL6 mice ($n = 6$) were infected with a 1:1 mixture of the indicated *S. Typhimurium* strains and the competitive index in colon contents determined 4 days after infection. Bars represent geometric means \pm standard errors (**d, e**). A Student's *t*-test was applied to determine statistical significance.

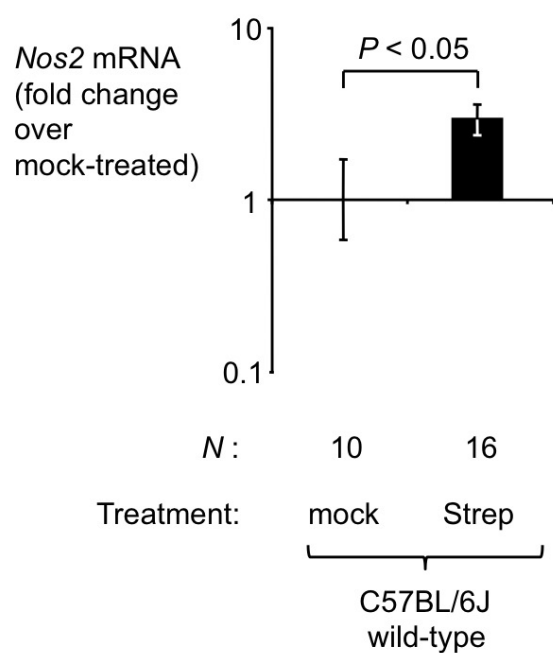


Extended Data Figure 2 | Evaluation of caecal inflammation in streptomycin-treated mice 4 days after *S. Typhimurium* infection. **a**, Streptomycin pre-treated mice were infected with the indicated strain mixtures and caecal histopathology was scored four days later for four mice per group. The criteria used for histopathology scoring

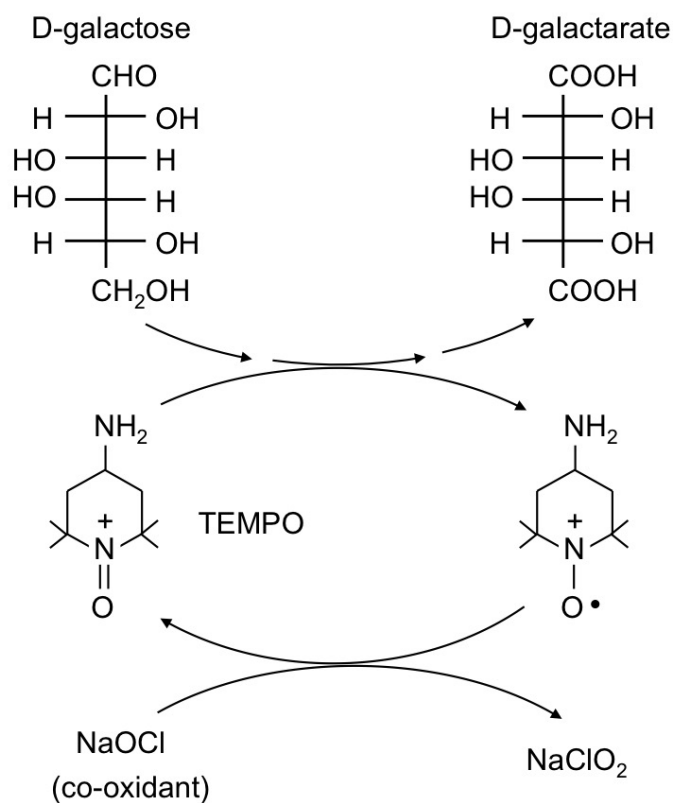
are listed in Extended Data Table 4. Each bar represents data from an individual animal. **b**, Representative images of haematoxylin and eosin (H&E)-stained caecal sections scored in **a**, along with an image from a mock-infected mouse for comparison. All images were taken at the same magnification. m, mucosa; s, submucosa; ml, muscle layer; lu, lumen.



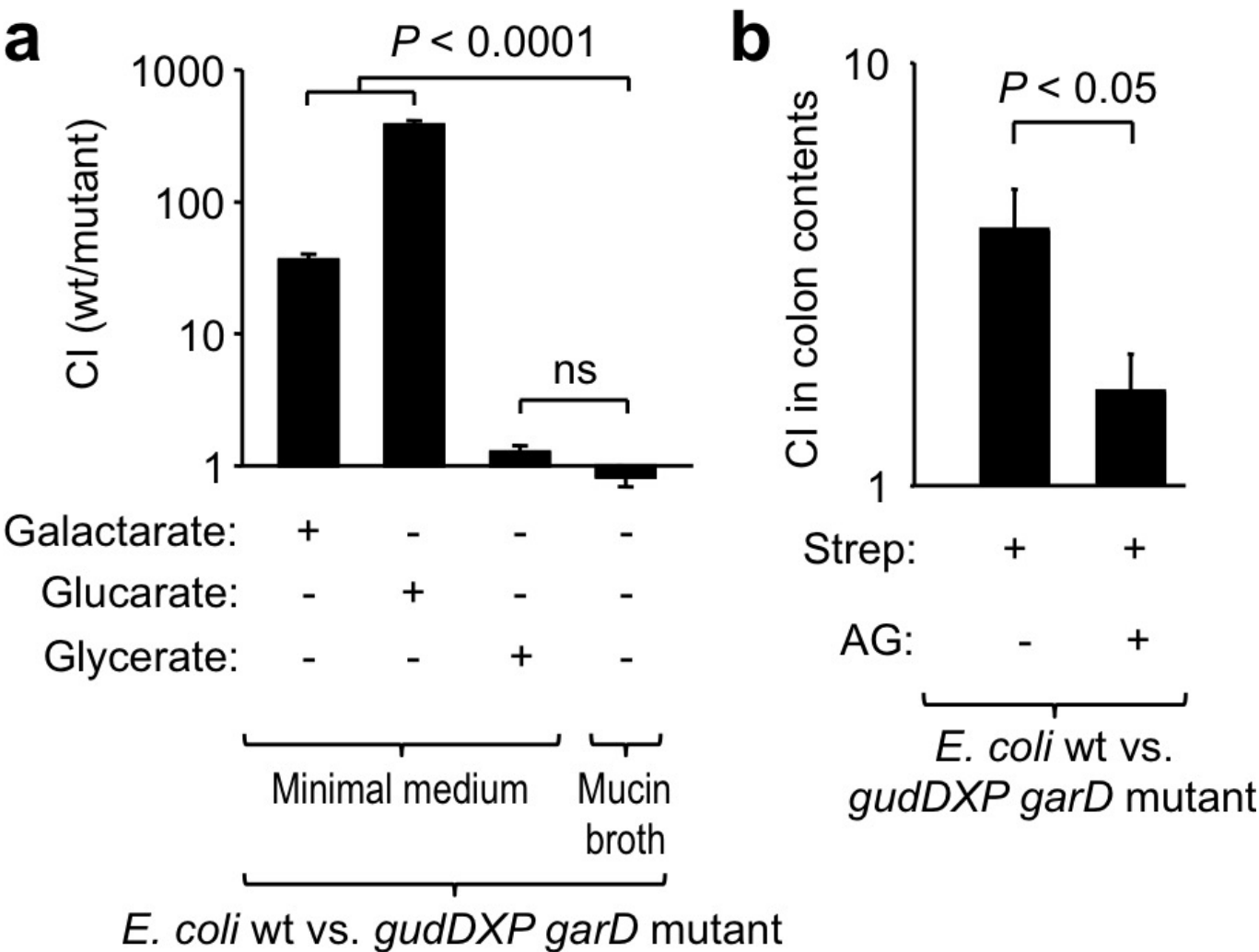
Extended Data Figure 3 | Detection of galactaric acid and glucaric acid by GC/MS. a, Representative GC elution profile of a caecal sample containing galactaric acid and glucaric acid (arrows). **b,** Representative single ion monitoring scan spectrum of galactaric acid and glucaric acid.

a

Extended Data Figure 4 | Elevated *Nos2* expression leads to nitrosyl radical-mediated oxidation of galactose. **a**, Expression levels of *Nos2* mRNA in RNA isolated from the caecal tip three days after mock-treatment (mock) or treatment of mice with streptomycin (Strep) was determined by quantitative real-time PCR. Bars represent geometric means \pm standard errors. A Student's *t*-test was applied to determine

b

statistical significance. **b**, Schematic of the oxidation of galactose to galactarate by TEMPO. 2,2,6,6-tetramethyl piperidine-1-oxyl (TEMPO) is a stable free nitrosyl radical that can oxidize terminal alcohol and aldehyde groups to carboxyl groups¹⁵. Consumption of TEMPO during the redox reaction is prevented by addition of a co-oxidant (NaOCl), which regenerates the nitrosyl radical.



Extended Data Figure 5 | Galactarate/glucarate fermentation by *E. coli*.
a, Minimal medium or mucin broth supplemented with the indicated carbon sources was inoculated with a 1:1 mixture of the *E. coli* wild type (wt) and a *gudDXP garD* mutant. CI, competitive index recovered after 24 h incubation in an anaerobic chamber. Growth was verified with 3 biological replicates. **b**, Streptomycin-treated C57BL6 mice ($n = 6$) were infected

with a 1:1 mixture of the indicated *E. coli* strains and received the iNOS inhibitor aminoguanidine (AG) or vehicle control. The competitive index in colon contents was determined four days after infection. Bars represent geometric means \pm standard errors (**a**, **b**). A Student's *t*-test was applied to determine statistical significance.

Extended Data Table 1 | Chart indicating scoring criteria for blinded examination of H&E-stained sections from the caecum

Score	Submucosal edema	Epithelial damage	Exudate	PMN infiltration*	Mononuclear cell infiltrate*
0	No changes	No changes	No changes	No changes (0-5)	No changes (0-5)
1	Detectable (<10%)	Desquamation	Slight accumulation	6-20	5-10
2	Mild (10%-20%)	Mild erosion	Mild accumulation	21-60	10-20
3	Moderate (20%-40%)	Marked erosion	Moderate accumulation	60-100	20-40
4	Marked (>40%)	Ulceration	Marked accumulation	>100	>40

*Number of cells per high-magnification microscopic field.

Extended Data Table 2 | Blinded histopathology scoring scheme

Combined score	Description
>8	Severe inflammation
6-8	Moderate inflammation
3-5	Mild inflammation
0-2	Normal

Extended Data Table 3 | Bacterial strains used in this study

Designation	Genotype	Reference
<i>E. coli</i>		
TOP10	F- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) Φ 80 <i>lacZ</i> Δ M15 <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ (<i>ara - leu</i>)7697 <i>galU</i> <i>galK</i> <i>rpsL</i> <i>endA1</i> <i>nupG</i>	Invitrogen
DH5a λ pir	F- <i>endA1</i> <i>hsdR17</i> (r-m+) <i>supE44</i> <i>thi-1</i> <i>recA1</i> <i>gyrA</i> <i>relA1</i> Δ (<i>lacZYA-argF</i>)U189 Φ 80 <i>lacZ</i> Δ M15 λ pir	Lab stock
S17-1 λ pir	<i>zxx::RP4</i> 2-(Tet ^R ::Mu) (Km ^R ::Tn7) λ pir	Lab stock
Nissle 1917	Wild-type strain (O6:K5:H1)	²⁴
FF441	Nissle 1917 <i>gudDXP</i> <i>garD</i>	This Study
<i>Salmonella</i>		
ATCC 14028	Wild-type isolate of <i>S. enterica</i> serovar Typhimurium	ATCC
IR715	Nalidixic acid-resistant derivative of ATCC14028	²⁵
CS019	ATCC 14028 <i>phoN::Tn10d-Cam</i>	²⁶
SPN487	IR715 Δ <i>invA</i> Δ <i>spiB</i>	²⁷
FF176	IR715 <i>phoN::Tn10d-Cam</i>	This Study
AJB715	IR715 <i>phoN::Km^R</i>	²⁸
FF183	IR715 <i>phoN::Tn10d-Cam</i> Δ <i>invA</i> Δ <i>spiB</i>	This Study
FF162	IR715 <i>gudTygcYgudDSTM2959::Km^R</i>	This Study
FF461	IR715 <i>phoN::Km^R</i> Δ <i>invA</i> Δ <i>spiB</i> <i>garD</i>	This Study
FF464	IR715 <i>phoN::Km^R</i> Δ <i>invA</i> Δ <i>spiB</i> <i>gudD</i>	This Study
FF217	IR715 Δ <i>invA</i> Δ <i>spiB</i> Δ <i>gudTygcYgudDSTM2959::Km^R</i>	This Study

References 24–28 are cited in this table.

Extended Data Table 4 | Plasmids used in this study

Designation	Relevant Characteristics	Reference
pCR2.1	Cloning vector, Carb ^R , Km ^R	Invitrogen
pRDH10	<i>ori</i> (R6K) <i>mobRP4 sacRB Tet^R Cm^R</i>	Lab stock
pBS34	pBluescript II KS+, KSAC cassette, Carb ^R , Km ^R	²⁹
pWSK29	Cloning vector, <i>ori</i> (pSC101) Carb ^R	Lab stock
pFF35	5' and 3' flanking regions of <i>gudTygcYgudDSTM2959</i> operon in pCR2.1, Carb ^R , Km ^R	This Study
pFF57	KSAC cassette flanked by up-/downstream regions of the <i>gudTygcYgudDSTM2959</i> operon in pRDH10; Cm ^R , Km ^R	This Study
pSW172	<i>ori</i> (R101) <i>repA101ts</i> Carb ^R	³⁰
pCAL61	<i>ori</i> (R101) Kan ^R Strep ^R	¹²
pCAL62	<i>ori</i> (R101) Carb ^R Strep ^R	¹²
pFF62	Up-/downstream regions of <i>gudD</i> from IR715 in pRDH10	This Study
pFF63	Up-/downstream regions of <i>garD</i> from IR715 in pRDH10	This Study
pFF64	Up-/downstream regions of the <i>gudDXP</i> operon from EcN in pRDH10	This Study
pFF65	Up-/downstream regions of <i>garD</i> from EcN in pRDH10	This Study
pGUDT	<i>gudTygcYgudDSTM2959</i> operon under the control of its native promoter in pWSK29; Carb ^R	This Study

References 29 and 30 are cited in this table.

Extended Data Table 5 | Primers used in this study

<i>Deletion of gudTygcYgudDSTM2959</i>		
Primer	Sequence*	
71	5'-GGATCCTCTGAACCGCTGCTAATGG-3'	
72	5'-TCTAGAGTTACGCTGAGTTGTAGG-3'	
73	5'-TCTAGAGTAGGGAATCAGAGATAACG-3'	
74	5'-GGATCCAGGGAGATACGCATAATGG-3'	
<i>Deletion of gudD in IR715</i>		
116	CACACCCGTCCTGTGCTCAACATGCACGATTCTG	
117	CGATCTCCCGATGTTTACCCATGTTG	
118	TAAACATCGGGAGATCGAACGTTTGTAAG	
119	GCGTCCGGCGTAGAGGAGCTTGACTGGGAACAG	
<i>Deletion of garD in IR715</i>		
112	CACACCCGTCCTGTGCTGAAATCAGAATGGGTC	
113	TCACAGGTCGGAATATGTTTCAGTCAGTTC	
114	CATATTCCGACCTGTGACCTGATCTATTCTG	
115	GCGTCCGGCGTAGAGATTGTCGCAAGGCTTCAC	
<i>Complementation of gudTygcYgudDSTM2959</i>		
92	5'-TCCTGCAGCCCGGGGCTTGCCTTGCCAGTAAGC-3'	
93	5'-TTAACGCACCATGCAAGGAC-3'	
94	5'-CTTGCATGGTGCGTTAAGTC-3'	
95	5'-CGCTCTAGAACTAGTGATCCGGCCTACAACCTCAGC-3'	
<i>Deletion of gudDXP operon in E. coli Nissle 1917</i>		
136	CACACCCGTCCTGTGCTGTGTTTATGCCGGATG	
137	CCGGTTCGTTCCCTGGCGATGTTTAC	
138	CCAGGGAACGAACCGGCAATAGAAAGC	
139	GCGTCCGGCGTAGAGTTTGCCTGGAGTCAAGCG	
<i>Deletion of garD in E. coli Nissle 1917</i>		
235	CACACCCGTCCTGTGTGGCCAACATCAAAATCAG	
236	CACCGGTGTTTCGGGTATTTTCGGTAG	
237	ACCCGAACCACCGGTGACCTGATTTTC	
238	GCGTCCGGCGTAGAGGCCAGCGACAAGTTTCTTTTC	
<i>Quantitative real-time RT-PCR</i>		
Organism	Target	Sequence
<i>Mus musculus</i>	<i>B2M</i>	5'-GGTCTTTCTGGTGCTTGTCTCA-3'
		5'-GTTTCGGCTTCCCATTTCTCC-3'
<i>Mus musculus</i>	<i>Nos2</i>	5'-TTGGGTCTTGTTCACCTCCACGG-3'
		5'-CCTCTTTCAGGTCACTTTGGTAGG-3'

*Restriction enzyme sites are underlined, overlapping sequences for Gibson Assembly are in bold.

Coordinating cardiomyocyte interactions to direct ventricular chamber morphogenesis

Peidong Han¹, Joshua Bloomekatz¹, Jie Ren¹, Ruilin Zhang¹, Jonathan D. Grinstein¹, Long Zhao², C. Geoffrey Burns², Caroline E. Burns², Ryan M. Anderson³ & Neil C. Chi^{1,4}

Many organs are composed of complex tissue walls that are structurally organized to optimize organ function. In particular, the ventricular myocardial wall of the heart comprises an outer compact layer that concentrically encircles the ridge-like inner trabecular layer. Although disruption in the morphogenesis of this myocardial wall can lead to various forms of congenital heart disease¹ and non-compaction cardiomyopathies², it remains unclear how embryonic cardiomyocytes assemble to form ventricular wall layers of appropriate spatial dimensions and myocardial mass. Here we use advanced genetic and imaging tools in zebrafish to reveal an interplay between myocardial Notch and *ErbB2* signalling that directs the spatial allocation of myocardial cells to their proper morphological positions in the ventricular wall. Although previous studies have shown that endocardial Notch signalling non-cell-autonomously promotes myocardial trabeculation through *ErbB2* and bone morphogenetic protein (BMP) signalling³, we discover that distinct ventricular cardiomyocyte clusters exhibit myocardial Notch activity that cell-autonomously inhibits *ErbB2* signalling and prevents cardiomyocyte sprouting and trabeculation. Myocardial-specific Notch inactivation leads to ventricles of reduced size and increased wall thickness because of excessive trabeculae, whereas widespread myocardial Notch activity results in ventricles of increased size with a single-cell-thick wall but no trabeculae. Notably, this myocardial Notch signalling is activated non-cell-autonomously by neighbouring *ErbB2*-activated cardiomyocytes that sprout and form nascent trabeculae. Thus, these findings support an interactive cellular feedback process that guides the assembly of cardiomyocytes to morphologically create the ventricular myocardial wall and more broadly provide insight into the cellular dynamics of how diverse cell lineages organize to create form.

The embryonic zebrafish heart comprises 200–300 cardiomyocytes when cardiac chambers form⁴, and thus provides an opportunity to interrogate in detail how individual cardiomyocytes organize to create the nascent structures of the vertebrate embryonic ventricular wall. As a result, previous zebrafish studies have shown that distinct cardiomyocytes extend from the embryonic ventricular wall into the lumen to develop cardiac trabeculae⁵, whereas others remain within this outer wall to create the primordial layer⁴. Yet, how these cardiomyocytes are selected to form the distinct myocardial layers of the ventricular wall remains to be fully elucidated.

Because of the role of Notch signalling in regulating cell–cell interactions^{6,7}, we examined its dynamic activation during zebrafish embryonic ventricular morphogenesis using the *Tg(Tp1:d2GFP)* Notch reporter line, which expresses a destabilized green fluorescent protein upon Notch activation⁸ (Fig. 1 and Extended Data Fig. 1). As previously reported⁹, we observed Notch signalling first in the ventricular endocardium at 24 hours post-fertilization (hpf), which then

becomes restricted to the atrioventricular (AV) and outflow tract (OFT) endocardium at 48 hpf (Extended Data Fig. 1a–l). From 72 to 96 hpf when cardiac trabeculation initiates^{5,10,11}, a subset of ventricular cardiomyocytes begins to express Notch-activated *Tp1:d2GFP* and remains

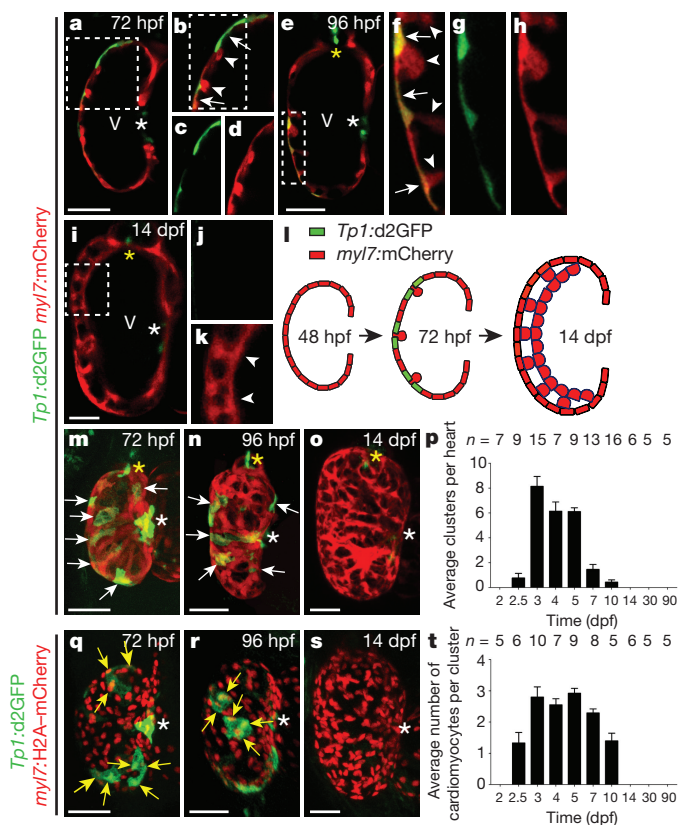


Figure 1 | Notch signalling is dynamically activated in distinct myocardial clusters during cardiac morphogenesis. Cardiac ventricles at 72 hpf, 96 hpf, and 14 dpf expressing (a–k, m–o) *Tp1:d2GFP*; *myl7:mCherry* or (q–s) *Tp1:d2GFP*; *myl7:H2A-mCherry*. a–k, Confocal slices; m–o, q–s, three-dimensional reconstructions. b, c–d, f–h, j–k, Magnifications of boxed areas in a, b, e, and i, respectively. Images c and d, g and h, and j and k are single channels from b, f, and i merged images, respectively. l, Schematic of myocardial Notch signalling. p, t, Quantification of (p) myocardial *Tp1:d2GFP*⁺ clusters and (t) cardiomyocytes per *Tp1:d2GFP*⁺ cluster. n, Number of embryos analysed per stage. White arrows, *Tp1:d2GFP*⁺ cardiomyocytes; white arrowheads, trabeculating cardiomyocytes; yellow arrows, cardiomyocytes in *Tp1:d2GFP*⁺ clusters. White and yellow asterisks, AV and OFT. Mean ± s.e.m. Scale bar, 25 μm.

¹Department of Medicine, Division of Cardiology, University of California, San Diego, La Jolla, California 92093, USA. ²Cardiovascular Research Center, Division of Cardiology, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts 02129, USA. ³Center for Diabetes and Metabolic Diseases, Department of Pediatrics and Department of Cellular and Integrative Physiology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ⁴Institute of Genomic Medicine, University of California, San Diego, La Jolla, California 92093, USA.

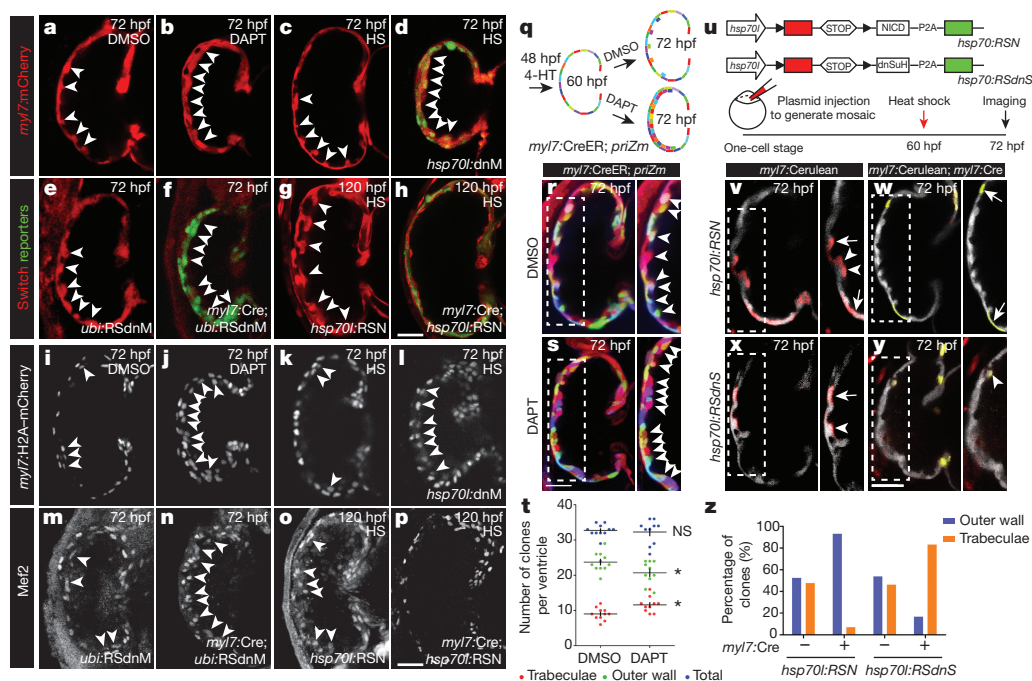


Figure 2 | Myocardial Notch signalling cell-autonomously regulates cardiomyocyte segregation between ventricular wall layers. Inhibiting Notch signalling by (b, j) DAPT treatment, (d, l) global dnMAML-GFP (*hsp70l:dnM*), or (f, n) myocardial-specific dnMAML expression (*myl7:Cre; ubi:RSdnM*) leads to excessive trabeculation at 72 hpf, whereas (h, p) myocardial-specific constitutive Notch activation via NICD expression (*myl7:Cre; hsp70l:RSN*) diminishes trabeculation at 120 hpf. a, c, e, g, i, k, m, o, Respective controls for each condition. a–p, For quantification see Extended Data Fig. 5. q, Myocardial *priZm* (brainbow) clonal studies. r, s, The 72 hpf *myl7:CreER; priZm* myocardial clones treated with DMSO or DAPT at 60 hpf. t, Although DMSO- and DAPT-treated ventricles display a similar overall number of myocardial clones (blue) ($n = 10$ and 11 embryos), DAPT-treated ventricles exhibit more

clones in trabeculae (red) and fewer in the outer ventricular wall (green), compared with control. Crosses, mean and s.e.m. * $P < 0.05$, by Student's *t*-test. NS, not significant. u, Notch-altering mosaic cardiomyocyte studies. w, Constitutively activated Notch cardiomyocytes expressing NICD-P2A-Emerald are primarily located on the ventricular outer wall ($n = 13/14$ clones, Fisher's exact test, $P < 0.05$); whereas (y) Notch-inhibited cardiomyocytes expressing dnSuH-P2A-Emerald are mainly found in trabeculae ($n = 15/18$ clones, Fisher's exact test, $P < 0.05$). v, x, In controls lacking *Tg(myl7:cre)*, mCherry⁺ cardiomyocytes are distributed equally between both layers ($n = 11/21$ and 14/26 clones in the outer wall). z, Quantitative analysis of v–y. Insets are magnifications of boxed areas. Arrowheads and arrows, trabeculae and outer wall cardiomyocytes. HS, heat shock. Scale bar, 25 μ m.

in the ventricular outer wall (Fig. 1a–h, arrows, and Extended Data Fig. 1m–o, arrows), whereas ventricular cardiomyocytes extending to form cardiac trabeculae fail to express *Tp1:d2GFP* (Fig. 1a–h arrowheads). These *Tp1:d2GFP*⁺ cardiomyocytes are frequently adjacent to sprouting *Tp1:d2GFP*[−] cardiomyocytes (Fig. 1a–h) and form clusters of two or three cardiomyocytes across the surface of the ventricular myocardial wall (Fig. 1m–t), which quantitatively correlate with the number of emerging cardiac trabeculae (Extended Data Fig. 1p–r). However, after the heart has established cardiac trabeculae, these *Tp1:d2GFP*⁺ myocardial clusters progressively decrease and are no longer observed by 14 days post-fertilization (dpf), despite the presence of *Tp1:d2GFP*⁺ AV and OFT endocardial cells (Fig. 1i–l, o, p, s, t and Extended Data Fig. 1s, t).

Using the *Tg(Tp1:eGFP)* Notch reporter line¹², which expresses a more stable fluorescent protein than *Tg(Tp1:d2GFP)* (Extended Data Fig. 2a–h), we observed that *Tp1:eGFP*⁺ cardiomyocytes remain present in the ventricular outer wall until 30–45 dpf because of the eGFP perdurance, and then become the ventricular primordial layer by 60–90 dpf when the ventricular cortical layer of the adult zebrafish heart forms⁴ (Extended Data Fig. 3a–n). Unlike ventricular trabecular and cortical cardiomyocytes, these *Tp1:eGFP*⁺ ventricular primordial cardiomyocytes fail to display organized sarcomeres by α -actinin immunostaining, are surrounded by extensive wheat germ agglutinin stained extracellular matrix and exhibit a thin cellular morphology as previously reported⁴ (Extended Data Fig. 3o–t). Overall, these data suggest that early myocardial Notch signalling may determine which ventricular cardiomyocytes remain in the embryonic ventricular outer wall to subsequently become the distinctive ventricular primordial layer of the adult zebrafish heart.

We next investigated the role of Notch signalling in the endocardium and myocardium during ventricular morphogenesis through selectively perturbing Notch signalling at specific cardiac developmental stages. Treating zebrafish embryos with DAPT, which effectively decreases *Tp1:d2GFP* Notch reporter expression and inhibits Notch signalling (Extended Data Fig. 2), from 20 to 48 hpf when Notch signalling is activated in the endocardium, reduces cardiac trabeculation (Extended Data Fig. 2q–s) as previously described³; however, treating from 60 to 72 hpf, when Notch signalling is present in the ventricular myocardium, results in increased trabeculae formation (Fig. 2a, b, i, j). Consistent with these results, BMP signalling, which is activated in ventricular trabeculae³, is also increased in similarly DAPT-treated zebrafish embryos from 60 to 72 hpf (Extended Data Fig. 4a–f). Furthermore, heat-shocking *Tg(hsp70l:dnMAML-GFP)*¹³ (abbreviated as *hsp70l:dnM*) embryos from 60 to 72 hpf, which induces dominant negative Mastermind-like (dnMAML) expression to block downstream Notch signalling, results in similar excessive trabeculation (Fig. 2c, d, k, l).

To explore whether Notch signalling functions in a cardiomyocyte-specific manner to directly guide myocardial cell fate position within the ventricle, we employed a myocardial-specific Cre (Extended Data Fig. 5a–d) strategy in combination with *Tg(ubi:loxP-mKate2-STOP-loxP-dnMAML-GFP)* or *Tg(hsp70l:loxP-mCherry-STOP-loxP-NICD-P2A-Emerald)*¹⁴ 'switch lines' (abbreviated as *ubi:RSdnM* and *hsp70l:RSN*) to inhibit or activate Notch signalling in cardiomyocytes, respectively. As observed in DAPT-treated and heat-shocked *Tg(hsp70l:dnM)* zebrafish from 60 to 72 hpf, *Tg(myl7:Cre; ubi:RSdnM)* zebrafish display excessive cardiac trabeculation due to inhibition of myocardial Notch signalling (Fig. 2e, f, m, n). Conversely,

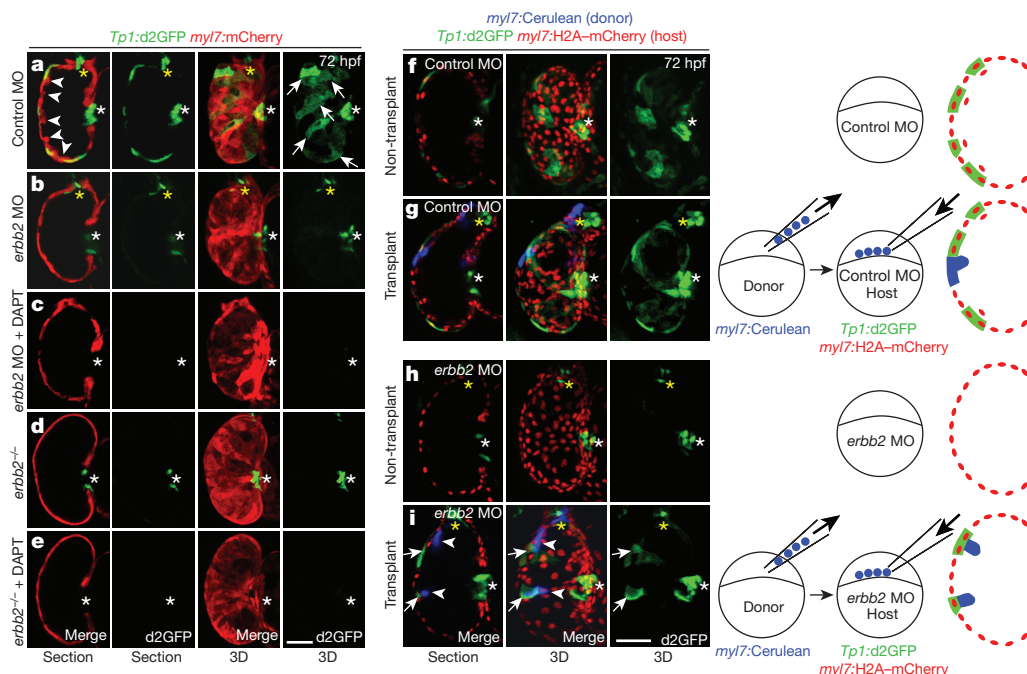


Figure 3 | Myocardial Erbb2 signalling non-cell-autonomously activates Notch signalling in neighbouring cardiomyocytes. Compared with (a) controls ($n = 0/15$ embryos), *Tp1:d2GFP; myl7:mCherry* (b) *erbb2* MO ($n = 10/12$) and (d) *erbb2*^{-/-} mutants ($n = 10/10$) display reduced trabeculae and myocardial Notch signalling. c, e, DAPT treatment at 60 hpf cannot rescue these myocardial defects, but can diminish AV and OFT endocardial Notch signalling (asterisks) ($n = 15/17$, $17/17$ embryos, respectively). f–i, Blastomere transplantation studies. Compared

with (g) control MO hosts ($n = 12$ embryos), (i) a greater percentage of donor *Tg(myl7:Cre; Cerulean)* wild-type cardiomyocytes are located in the trabeculae of *erbb2* MO *Tg(Tp1:d2GFP; myl7:H2A-mCherry)* hosts ($n = 10$). In contrast to (h) non-transplanted *erbb2* MO hearts ($n = 16$), (i) transplanted donor *Tg(myl7:Cre; Cerulean)* wild-type cardiomyocytes (arrowheads) can activate myocardial Notch activity (*Tp1:d2GFP*) in neighbouring *erbb2* MO *Tg(Tp1:d2GFP; myl7:H2A-mCherry)* host cardiomyocytes (arrows, $n = 10$). Scale bar, 25 μ m.

heat-shocking *Tg(myl7:Cre; hsp70l:RSN)* zebrafish, which induces myocardial Notch-intracellular domain (NICD) expression, between 60 and 120 hpf leads to cardiac ventricles without significant trabeculae because of constitutively activated Notch signalling throughout the myocardium (Fig. 2g, h, o, p). Moreover, constitutive myocardial Notch activation at later time points (80, 96, and 120 hpf to 7 dpf) prevents trabeculae from further sprouting and/or extending (Extended Data Fig. 6a–g); however, trabeculae continue to develop after the cessation of this myocardial Notch activity, but fail to recover to wild-type levels (Extended Data Fig. 6h, i).

In line with these findings, we discovered that Notch inhibition results in smaller ventricular areas (Fig. 2a–f and Extended Data Fig. 5e) and thicker ventricular myocardial walls (Fig. 2a–f and Extended Data Fig. 5f) due to increased cardiomyocytes within the trabecular layer (approximately two or three cells thick) (Fig. 2i–n, Extended Data Fig. 5g), whereas Notch activation gives rise to larger ventricular areas (Fig. 2g, h, Extended Data Fig. 5e) and thinner ventricular myocardial walls (Fig. 2g, h and Extended Data Fig. 5f) that are about one cell thick with no apparent trabecular cardiomyocytes (Fig. 2o, p and Extended Data Fig. 5g). Although these hearts do not exhibit a significant difference in overall cardiomyocyte numbers compared with control hearts (Extended Data Fig. 5h–p), we did discover that Notch inhibition promotes the redistribution of N-cadherin away from intercellular contacts whereas Notch activation prevents this reorganization (Extended Data Fig. 7), suggesting that myocardial Notch signalling may control ventricular size and wall thickness through regulating the allocation of cardiomyocytes between the ventricular wall layers via cell–cell contacts. To further investigate this possibility, we monitored the fate of individual genetically labelled cardiomyocytes using a myocardial specific Brainbow system *Tg(myl7:CreER; priZm)*⁴ (Fig. 2q–t). After confirming that adjacent cardiomyocytes were consistently labelled with different colours at 60 hpf before trabeculation (Extended Data Fig. 8), we treated zebrafish embryos with DAPT or

dimethylsulfoxide (DMSO) from 60 to 72 hpf. DAPT treatment leads to increased numbers of trabeculating clones and conversely decreased numbers of non-trabeculating clones compared with DMSO-treated hearts; however, the total number of ventricular cardiomyocyte clones is not significantly different (Fig. 2r–t), further supporting the idea that Notch signalling segregates individual cardiomyocyte clones between the ventricular outer wall and inner trabecular layers.

To examine whether Notch signalling acts cell-autonomously to control cardiomyocyte sprouting, we perturbed Notch signalling in individual cardiomyocytes during trabeculation by injecting *hsp70l:loxP-mCherry-STOP-loxP-NICD-P2A-Emerald* (*hsp70l:RSN*, Notch activating) or *hsp70l:loxP-mCherry-STOP-loxP-dnSuH-P2A-Emerald*¹⁵ (*hsp70l:RSdnS*, dominant negative Suppressor of Hairless/Notch repressing) switch plasmids into *Tg(myl7:Cre)*¹⁶; *Tg(myl7:Cre)* zebrafish embryos (Fig. 2u). Heat-shocking these injected fish from 60 to 72 hpf resulted in most Notch-activated NICD-P2A-Emerald⁺ cardiomyocytes remaining in the ventricular outer myocardial wall (Fig. 2w, z), whereas Notch-inhibited dnSuH-P2A-Emerald⁺ cardiomyocytes reside primarily in trabeculae (Fig. 2y, z). Heat-shocking injected control fish lacking *Tg(myl7:Cre)* generated mCherry⁺ cardiomyocytes that were distributed equally between both myocardial layers (Fig. 2v, x, z), altogether revealing a myocardial cell-autonomous role for Notch signalling.

Because Neuregulin/Erbb2 and BMP10 signalling can promote cardiac trabeculation^{3,10,17}, we investigated whether myocardial Notch may cross-talk with these signalling pathways to regulate cardiomyocyte selection between the ventricular wall layers. Inhibiting Erbb2 signalling with AG1478 treatment¹⁰ from 60 to 72 hpf prevents cardiac trabeculation and expression of *Tp1:d2GFP* in cardiomyocytes, although *Tp1:d2GFP* remains expressed in AV and OFT endocardial cells (Extended Data Fig. 9a, b). Consistent with these findings, both *erbb2* morpholino (MO) knockdown and *erbb2*^{-/-} mutant (*erbb2*^{st50}) *Tg(Tp1:d2GFP)* embryos, which exhibit similar trabecular defects^{10,18},

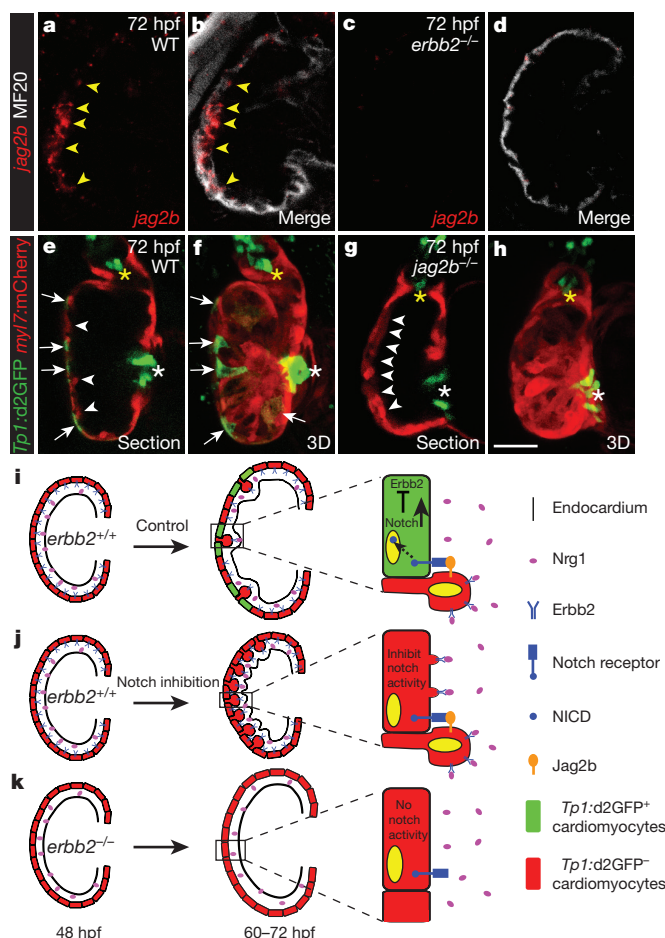


Figure 4 | The Notch ligand Jag2b mediates cooperative interactions between cardiomyocytes. **a–d**, Ventricular myocardial (MF20⁺) *jag2b* is expressed in **(a, b)** wild-type (WT) ($n = 6/6$ embryos) but not **(c, d)** *erb2*^{−/−} mutant hearts ($n = 0/5$). **e–h**, Compared with **(e, f)** WT controls ($n = 0/10$), **(g, h)** *Tg(Tp1:d2GFP; myl7:mCherry) jag2b*^{−/−} mutants exhibit increased trabeculation and reduced myocardial Notch signalling at 72 hpf ($n = 8/8$). Yellow arrowheads, *jag2b*⁺ cardiomyocytes; white arrowheads, trabeculae; white arrows, *Tp1:d2GFP*⁺ cardiomyocytes; white and yellow asterisks, AV and OFT. Scale bar, 25 μ m. Myocardial Notch signalling model: **(i)** endocardial Neuregulin/Nrg1 activates myocardial Erb2 signalling, which cell-autonomously triggers myocardial sprouting and Jag2b expression (60–72 hpf). Jag2b activates Notch signalling in neighbouring cardiomyocytes, which cell-autonomously inhibits *erb2* expression and trabeculae formation (magnified area). **j**, Inhibiting Notch signalling allows all cardiomyocytes to express *erb2*, respond to Neuregulin, and sprout and form trabeculae. **k**, Blocking Erb2 signalling prevents trabeculation, Jag2b expression, and Notch activation in neighbouring cardiomyocytes.

also fail to display Notch reporter expression in the myocardium (Fig. 3a, b, d), supporting a requirement for Erb2 signalling in the initiation of trabeculation and the activation of myocardial Notch signalling. Notably, neither DAPT treatment nor heat-shock induction of dnMAML between 60 and 72 hpf, which alone increases trabeculation (Fig. 2), could rescue the relative lack of cardiac trabeculae in zebrafish with loss of Erb2 function (Fig. 3c, e and Extended Data Fig. 9b–e). In contrast to the Erb2 loss of function findings, embryos treated with Dorsomorphin¹⁹ from 60 to 72 hpf to inhibit BMP signalling still form cardiac trabeculae and express *Tp1:d2GFP* in the outer myocardial wall (Extended Data Fig. 4m–p) despite the abrogation of the BMP-reporter signal (*BRE:d2GFP*) in the trabecular layer (Extended Data Fig. 4j–l). However, by 7 dpf, these hearts display aberrant and stunted cardiac trabeculae compared with DMSO-treated fish (Extended Data Fig. 4q–s), corroborating a

requirement for BMP signalling in the maintenance but not the initiation of cardiac trabeculation²⁰.

To explore whether Notch activation negatively regulates Erb2 signalling to prevent trabeculae formation, we examined *erb2* expression in 72 hpf *Tp1:d2GFP* hearts and discovered that *erb2* is expressed in many ventricular cardiomyocytes but diminished in *Tp1:d2GFP*⁺ cardiomyocytes (Extended Data Fig. 9f–j). In support of these findings, constitutive myocardial Notch activation by heat-shocking *Tg(myl7:Cre; hsp70l:RSN)* fish between 60 and 120 hpf results in the dramatic reduction of *erb2* myocardial expression (Extended Data Fig. 9k, l, o, p). In contrast, Notch-inhibited hearts treated with DAPT from 60 to 72 hpf exhibit increased *erb2* myocardial expression (Extended Data Fig. 9m, n, q, r). Thus, myocardial Notch signalling may block Neuregulin/Erb2 signalling by downregulating *erb2* expression to inhibit cardiomyocyte sprouting.

Since Notch signalling has been shown to mediate cell fate position through lateral inhibition mechanisms^{6,7,21,22}, we investigated whether Erb2 signalling non-cell-autonomously activates myocardial Notch signalling in neighbouring cardiomyocytes. Thus, we created mosaic embryos by transplanting *Tg(myl7:Cerulean)* wild-type donor blastomeres into *erb2* or control MO injected *Tg(Tp1:d2GFP; Tg(myl7:H2A-mCherry))*²³ host embryos and assessed the ability of wild-type donor cells to contribute to the ventricular wall layers and activate myocardial Notch signalling. As previously reported¹⁰, a greater percentage of donor-derived wild-type cardiomyocytes is present in the trabeculae of *erb2* knockdown embryos compared with control embryos (compare Fig. 3i with Fig. 3g; Extended Data Fig. 10a). Although non-transplanted *erb2* knockdown hearts fail to exhibit myocardial Notch activity (Fig. 3f, h), transplanted *erb2* knockdown host hearts containing wild-type donor myocardial cells (*myl7:Cerulean*⁺) can activate myocardial *Tp1:d2GFP* expression (Fig. 3i and Extended Data Fig. 10b). Upon closer inspection, these host *erb2* knockdown *Tp1:d2GFP*⁺ cardiomyocytes (Fig. 3i, arrows) appear adjacent to donor wild-type *myl7:Cerulean*⁺ cardiomyocytes (Fig. 3i, arrowheads, and Extended Data Fig. 10c), supporting a role for Erb2-responsive cardiomyocytes in activating Notch signalling in neighbouring cardiomyocytes.

On the basis of these results, we searched for potential Notch ligands mediating the activation of Notch signalling in neighbouring cardiomyocytes and discovered that *jag2b* is expressed in select ventricular cardiomyocytes at 72 hpf when myocardial Notch signalling is activated (Fig. 4a, b). This ventricular myocardial *jag2b* expression is reduced in *erb2*^{−/−} mutant hearts (Fig. 4c, d), suggesting that Erb2 signalling may activate Notch signalling in neighbouring cardiomyocytes through *jag2b*. In support of this possibility, we discovered that *jag2b*^{−/−} mutant hearts exhibit not only increased trabeculation as observed in Notch-inhibited hearts but also reduced *Tp1:d2GFP* Notch reporter activity in the ventricular myocardium but not in the AV or OFT endocardium (Fig. 4e–h). Together these data support a model in which myocardial Erb2 signalling non-cell-autonomously activates Notch signalling in neighbouring ventricular outer-wall cardiomyocytes through Jag2b, which in turn leads to the reduction of *erb2* expression, subsequent inhibition of Erb2 signalling, and suppression of cardiomyocyte sprouting (Fig. 4i–k).

Overall, these findings reveal a molecular mechanism whereby Notch and Erb2 signalling coordinates social cellular interactions between cardiomyocytes that determine their morphological fate within the ventricular wall. Although previous studies have suggested that Notch signalling may be activated in the myocardium^{24–27}, our zebrafish studies illuminate the precise role of myocardial Notch activity in forming the ventricular wall. Similar to the receptor tyrosine kinase (RTK)-Notch lateral inhibition signalling mechanisms that regulate epithelial tip and stalk cell formation during branching morphogenesis^{6,7}, myocardial Notch acts in concert with the RTK Erb2 to segregate embryonic cardiomyocytes into two functionally distinct classes of cells: (1) sprouting cardiomyocytes that respond to Neuregulin via

ErbB2 and (2) non-sprouting Notch-activated cardiomyocytes, in which Notch signalling inhibits *erbb2* expression. These roles appear not to be pre-specified, but rather are determined by social interactions between cardiomyocytes. Furthermore, recent studies have reported human Notch genetic variants linked to a wide spectrum of congenital heart diseases including non-compaction cardiomyopathies^{28,29}, which exhibit similar severe ventricular wall defects to those observed in our Notch studies. More broadly, our studies support a conserved role for intercellular cross-talk between RTKs and Notch signalling for allocating cells within organ substructures and might be particularly relevant in developing strategies for human pluripotent stem-cell tissue-specific developmental and disease modelling or regenerative therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 September 2015; accepted 5 May 2016.

- Fahed, A. C., Gelb, B. D., Seidman, J. G. & Seidman, C. E. Genetics of congenital heart disease: the glass half empty. *Circ. Res.* **112**, 707–720 (2013).
- Zemrak, F. *et al.* The relationship of left ventricular trabeculation to ventricular function and structure over a 9.5-year follow-up: the MESA study. *J. Am. Coll. Cardiol.* **64**, 1971–1980 (2014).
- Grego-Bessa, J. *et al.* Notch signaling is essential for ventricular chamber development. *Dev. Cell* **12**, 415–429 (2007).
- Gupta, V. & Poss, K. D. Clonally dominant cardiomyocytes direct heart morphogenesis. *Nature* **484**, 479–484 (2012).
- Staudt, D. W. *et al.* High-resolution imaging of cardiomyocyte behavior reveals two distinct steps in ventricular trabeculation. *Development* **141**, 585–593 (2014).
- Ghabrial, A. S. & Krasnow, M. A. Social interactions among epithelial cells during tracheal branching morphogenesis. *Nature* **441**, 746–749 (2006).
- Siekman, A. F. & Lawson, N. D. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445**, 781–784 (2007).
- Clark, B. S. *et al.* Loss of Lgl1 in retinal neuroepithelia reveals links between apical domain size, Notch activity and neurogenesis. *Development* **139**, 1599–1610 (2012).
- Samsa, L. A. *et al.* Cardiac contraction activates endocardial Notch signaling to modulate chamber maturation in zebrafish. *Development* **142**, 4080–4091 (2015).
- Liu, J. *et al.* A dual role for ErbB2 signaling in cardiac trabeculation. *Development* **137**, 3867–3875 (2010).
- Ong, L. L., Kim, N., Mima, T., Cohen-Gould, L. & Mikawa, T. Trabecular myocytes of the embryonic heart require N-cadherin for migratory unit identity. *Dev. Biol.* **193**, 1–9 (1998).
- Parsons, M. J. *et al.* Notch-responsive cells initiate the secondary transition in larval zebrafish pancreas. *Mech. Dev.* **126**, 898–912 (2009).
- Zhao, L. *et al.* Notch signaling regulates cardiomyocyte proliferation during zebrafish heart regeneration. *Proc. Natl Acad. Sci. USA* **111**, 1403–1408 (2014).
- Anderson, R. M. *et al.* Hepatocyte growth factor signaling in intrapancreatic ductal cells drives pancreatic morphogenesis. *PLoS Genet.* **9**, e1003650 (2013).
- Ninov, N., Borius, M. & Stainier, D. Y. Different levels of Notch signaling regulate quiescence, renewal and differentiation in pancreatic endocrine progenitors. *Development* **139**, 1557–1567 (2012).
- Mathews, E. S. *et al.* Mutation of 3-hydroxy-3-methylglutaryl CoA synthase I reveals requirements for isoprenoid and cholesterol synthesis in oligodendrocyte migration arrest, axon wrapping, and myelin gene expression. *J. Neurosci.* **34**, 3402–3412 (2014).
- Peshkovsky, C., Totong, R. & Yelon, D. Dependence of cardiac trabeculation on neuregulin signaling and blood flow in zebrafish. *Dev. Dyn.* **240**, 446–456 (2011).
- Lyons, D. A. *et al.* *erbb3* and *erbb2* are essential for Schwann cell migration and myelination in zebrafish. *Curr. Biol.* **15**, 513–524 (2005).
- Colley, R. F. & Link, B. A. Dynamic smad-mediated BMP signaling revealed through transgenic zebrafish. *Dev. Dyn.* **240**, 712–722 (2011).
- Chen, H. *et al.* BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* **131**, 2219–2231 (2004).
- Sternberg, P. W. Lateral inhibition during vulval induction in *Caenorhabditis elegans*. *Nature* **335**, 551–554 (1988).
- Heitzler, P. & Simpson, P. The choice of cell fate in the epidermis of *Drosophila*. *Cell* **64**, 1083–1092 (1991).
- Schumacher, J. A., Bloomekatz, J., Garavito-Aguilar, Z. V. & Yelon, D. *tal1* regulates the formation of intercellular junctions and the maintenance of identity in the endocardium. *Dev. Biol.* **383**, 214–226 (2013).
- Rentschler, S. *et al.* Notch signaling regulates murine atrioventricular conduction and the formation of accessory pathways. *J. Clin. Invest.* **121**, 525–533 (2011).
- Zhao, C. *et al.* Numb family proteins are essential for cardiac morphogenesis and progenitor differentiation. *Development* **141**, 281–295 (2014).
- Yang, J. *et al.* Inhibition of Notch2 by Numb/Numbl controls myocardial compaction in the heart. *Cardiovasc. Res.* **96**, 276–285 (2012).
- Leimeister, C., Externbrink, A., Klamt, B. & Gessler, M. Hey genes: a novel subfamily of hairy- and Enhancer of split related genes specifically expressed during mouse embryogenesis. *Mech. Dev.* **85**, 173–177 (1999).
- Li, L. *et al.* Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nature Genet.* **16**, 243–251 (1997).
- Luxán, G. *et al.* Mutations in the NOTCH pathway regulator MIB1 cause left ventricular noncompaction cardiomyopathy. *Nature Med.* **19**, 193–201 (2013).

Acknowledgements We thank N. Tedeschi for fish care; B. Le for experimental assistance; S. Evans, D. Yelon, and Chi laboratory members for comments on the manuscript; N. Ninov and D. Stainier for plasmids; B. Link for the d2GFP BMP and d2GFP Notch reporter lines; N. Lawson for the eGFP Notch reporter line; B. Appel for the myocardial Cerulean line; K. Poss for the myocardial CreER and Brainbow/priZm lines; and W. Talbot for the *erbb2* mutant. This work was supported in part by grants from American Heart Association (14POST20380738) to L.Z.; the March of Dimes (1-FY14-327) to R.A.M.; the NIH/NHLBI (5R01HL127067) to C.G.B. and C.E.B.; and the National Institutes of Health to N.C.C.

Author Contributions P.H. and N.C.C. conceived the project and the design of the experimental strategy. P.H., J.R., J.B., R.Z., and J.D.G. conducted experiments. L.Z. generated the *ubi:RSdnM* transgenic line. P.H. and N.C.C. generated and characterized the *myl7:Cre* transgenic line. C.E.B., C.G.B., and R.A.M. provided key reagents. P.H., J.B. and N.C.C. prepared the manuscript.

Author Information Reprints and permission information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.C.C. (nchi@ucsd.edu).

Reviewer Information Nature thanks B. G. Bruneau and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Zebrafish husbandry and strains. Zebrafish (*Danio rerio*) were raised under standard laboratory conditions at 28 °C. All animal work was approved by the University of California at San Diego Institutional Animal Care and Use Committee. The following established transgenic and mutant lines were used: *Tg(EPV:Tp1-Mmu.Hbb:d2GFP)^{mw43}* (ref. 8) abbreviated as *Tg(Tp1:d2GFP)*; *Tg(EPV:Tp1-Mmu.Hbb:eGFP)^{um14}* (ref. 12) abbreviated as *Tg(Tp1:eGFP)*; *Tg(BRE-AAVmlp:d2GFP)^{mw30}* (ref. 19) abbreviated as *Tg(BRE:d2GFP)*; *Tg(hsp70l:dnMAML-GFP)^{b10}* (ref. 13) abbreviated as *Tg(hsp70l:dnM)*; *Tg(kdrl:ras-mCherry)^{s896}* (ref. 30); *Tg(myl7:H2A-mCherry)^{sd12}* (ref. 23); *Tg(myl7:mCherry)^{sd7}* (ref. 31); *Tg(myl7:Cre)^{co19}* (ref. 16); *Tg(myl7:eGFP-HRAS)^{s883}* (ref. 32) abbreviated as *Tg(myl7:ras-eGFP)*; *Tg(myl7:CreER)^{pd10}* (ref. 33); *Tg(β-act2:Brainbow1.0L)^{bd49}* (ref. 4) abbreviated as *Tg(priZm)*; *Tg(hsp70l:loxP-mCherry-STOP-loxP-NICD-P2A-Emerald)^{s961}* (ref. 14) abbreviated as *Tg(hsp70l:RSN)*; *Tg(β-act2:loxP-DsRed-STOP-loxP-eGFP)^{s928}* (ref. 33) abbreviated as *Tg(β-act2:RSG)*; *erbb2^{st50}* (ref. 18), and *jag2b^{hu3425}* (ref. 34).

To generate the *Tg(myl7:Cre)^{sd38}* transgenic line, a 900-base-pair fragment of the *myl7* promoter³⁵ was cloned upstream of the *Cre* recombinase gene into a multi-cloning site flanked by I-SceI sites in the pBluescript-SK vector. Standard I-SceI meganuclease transgenesis³⁶ was used to create transgenic founders which were screened for myocardial *Cre* recombinase activity by crossing to the *Tg(β-act2:RSG)^{s928}* line. Three independent founders were identified, all with similar levels of *Cre* recombinase activity and matching the expression of *Tg(myl7:Cre)^{co19}* (Extended Data Fig. 5a–d). A single representative founder was propagated further.

To generate the *Tg(ubi:loxP-mKate2-STOP-loxP-dnMAML-GFP)^{b16}* strain, abbreviated as *Tg(ubi:RSdnM)*, gateway cloning technology (Life Technologies) was used to conduct an LR recombination reaction with the *pENTR5'-ubi³⁷*, *pME-loxP-mKate2-STOP-loxP*, *p3E-dnMAML-GFP* entry vectors and the *pDESTol2pA2* destination vector³⁸. The *pME-loxP-mKate2-STOP-loxP* entry vector was created by replacing the *AmCyan* complementary DNA in the *pME-loxP-AmCyan-STOP-loxP³⁹* vector with a complementary DNA encoding mKate2 (Evrogen) using In-Fusion HD cloning (Clontech Laboratories). The *p3E-dnMAML-GFP* entry vector was generated by conducting a BP recombination reaction between a PCR product encoding a fusion protein between dnMAML and GFP amplified from *pME-dnMAML-GFP¹³* and the Gateway Donor Vector *pDONR2P-P3*. Altogether the *ubi:loxP-mKate2-STOP-loxP-dnMAML-GFP* construct was co-injected with *Tol2* transposase mRNA³⁸ into one-cell stage embryos to generate independent founders which were screened for mKate2 and then GFP upon *Cre*-mediated recombination. Founders with both mKate2 and GFP were propagated further.

Embryonic immunofluorescence and live imaging studies. Wholemount immunofluorescence studies were performed as previously described⁴⁰, with the following modifications. After initial fixation, any pre-existing fluorescence was quenched by incubating embryos in 2 M HCl at 37 °C for 30 min and washing with double-distilled H₂O and phosphate buffer saline with 0.1% Tween-20 (PBST). The antibodies used were anti-Mef2/C-21 (rabbit, Santa Cruz Biotechnology, 1:100), anti-MHC/MF20 (mouse, Developmental Studies Hybridoma Bank, 1:100) or anti-N-cadherin (rabbit, GeneTex, 1:100) followed by anti-rabbit IgG-Alexa 488 (goat, Life Technologies 1:200).

For embryonic studies of Notch activity, embryos containing the Notch reporters *Tg(Tp1:d2GFP)^{mw43}* or *Tg(Tp1:eGFP)^{um14}* in combination with myocardial-expressed transgenes such as *Tg(myl7:mCherry)^{sd7}* or endothelial expressed transgenes such as *Tg(kdrl:ras-mCherry)^{s896}* were imaged live⁴⁰. The *Tp1* promoter used in these Notch reporter transgenics consists of 12 RPBj-binding sites and reports Notch activity throughout the embryo as previously published^{8,12}. These embryos were embedded in 1% low melting agarose (Lonza) in a coverslip bottom culture dish (MatTek) and cardiac contraction was arrested using Tricaine-S (Sigma MS-222) just before imaging.

To count the number of *Tp1:d2GFP⁺* clusters in each heart, we used three-dimensional reconstructions (Nikon NIS Elements software) from confocal stacks of *Tg(Tp1:d2GFP; myl7:mCherry)* embryos (5–16 hearts per stage). To count the number of cardiomyocytes per *Tp1:d2GFP⁺* cluster, *Tg(Tp1:d2GFP; myl7:H2A-mCherry)* embryos were used. Only cells expressing both H2A-mCherry and d2GFP were counted. We analysed five to ten hearts per stage to obtain the average number of cardiomyocytes in each *Tp1:d2GFP* cluster at each specified stage. Statistical analysis is described in the 'Image processing and statistical analysis' section below.

To assess the correlation between nascent trabeculae and *Tp1:d2GFP⁺* clusters at 72 hpf, trabeculae were identified in consecutive slices from a Z-stack and pseudo-coloured with magenta. A three-dimensional reconstruction was then used to generate the full view of the ventricle, where the number of nascent trabeculae and *Tp1:d2GFP⁺* clusters could be counted in each heart. These data

are represented in a scatter plot (Extended Data Fig. 1r) and used to overlay a linear regression line.

Adult immunofluorescence and imaging studies. Immunofluorescence studies were conducted on cryosections of adult zebrafish hearts. These hearts were cryoprotected, mounted, sectioned, and stained as performed previously⁴⁰. The following primary antibodies were used: anti-MHC/MF20 (mouse, Developmental Studies Hybridoma Bank, 1:100); anti-α-actinin (mouse, Diagnostic BioSystems, 1:100); anti-Raldh2 (rabbit, Abmart, 1:100); and anti-GFP (chicken, Aves Labs, 1:200). The following secondary antibodies were used: anti-mouse IgG-Alexa 405 (goat, Life technologies, 1:200), anti-mouse IgG-Alexa 594 (goat, Life Technologies, 1:200), anti-rabbit IgG-Alexa 568 (goat, Life Technologies, 1:200) and anti-chicken IgG-Alexa 488 (goat, Life Technologies, 1:200). Alexa Fluor 594-conjugated wheat germ agglutinin (Life Technologies, 50 µg ml⁻¹) was used to stain the extracellular matrix. DAPI (1 µg ml⁻¹) staining was used to identify nuclei. Notably, we discovered that eGFP from the *Tg(Tp1:eGFP)^{um14}* transgene perdured for a longer period in the ventricular outer myocardial wall (Extended Data Fig. 3) than d2GFP from the *Tg(Tp1:d2GFP)^{mw43}* transgene (Extended Data Fig. 1).

Notch signalling studies. Notch inhibition studies were performed using DAPT (a chemical inhibitor of γ-secretase) or dnMAML mis-expression (dominant negative mastermind-like 1). DAPT: zebrafish embryos were incubated in 100 µM DAPT (Sigma) or 0.1% DMSO alone (control) at specified developmental stages and time intervals and then quickly washed (two or three times) with egg water (60 µg ml⁻¹ Instant Ocean sea salts) for further analysis. The ability of 100 µM DAPT treatment to inhibit Notch signalling was validated by examining *Tp1:d2GFP* expression after DAPT treatment (Extended Data Fig. 2i–p). dnMAML: The *Tg(hsp70l:dnM)* was used to globally express dnMAML at specified time points. Heat-shock induction was conducted by placing *Tg(hsp70l:dnM)* or wild-type siblings into a 37 °C incubator for 30 min, followed by 3 min in a 42 °C water bath. Embryos were heat-shocked twice every 24 h to maintain the induction of dnMAML-GFP throughout the embryo. This protocol was highly efficient at inducing dnMAML-GFP expression and produced minimal lethality. To inhibit Notch signalling in cardiomyocytes only, the *Tg(ubi:loxP-mKate2-STOP-loxP-dnMAML-GFP)* line was crossed with the *Tg(myl7:Cre)* line to produce embryos which express dnMAML-GFP only in the myocardium. Induction of dnMAML-GFP was verified by examining GFP fluorescence 5–6 h after heat shock or *Cre*-mediated recombination¹³.

Notch activation was performed by expressing NICD. Heat shocking embryos containing both *Tg(hsp70l:loxP-mCherry-STOP-loxP-NICD-P2A-Emerald)¹⁴* and *Tg(myl7:Cre)* transgenes produced NICD-P2A-Emerald only in the myocardium. Heat shock was performed as described above.

Ventricular wall thickness was measured to quantify the effect of perturbing Notch signalling and was determined by drawing five representative lines perpendicular to the ventricular wall in a representative confocal slice. Thickness was measured as the distance along the line between the lateral and medial edge of the myocardial wall. All hearts were imaged in the same orientation and comparable confocal slices were chosen for analysis. Six hearts were measured for each condition.

To determine the effect of altering Notch signalling on cardiomyocyte cell numbers within the ventricular outer wall and trabecular layers, ventricular cardiomyocyte nuclei were counted from hearts exposed to specified experimental conditions using three-dimensional reconstructions of confocal slices from embryos with *myl7:H2A-mCherry* or from embryos stained with the Mef2 antibody. Mef2 immunostaining was used in embryos containing transgenes with fluorophores that overlapped with H2A-mCherry, such as *ubi:RSdnM* or *hsp70l:RSN*. For these analyses, the cells within the trabeculae could be separated from the ventricular outer wall using the post-image processing procedure described in the 'Image processing and statistical analysis' section below. Using this procedure, we calculated the number of cardiomyocyte nuclei in the total ventricle and the number of cardiomyocyte nuclei within the trabeculae. The number of cardiomyocyte nuclei in the ventricular outer wall was calculated by subtracting the number of cardiomyocytes in the trabeculae from the total.

Trabeculae area was measured from a confocal slice of a ventricle containing a cytoplasmic fluorophore such as *myl7:mCherry* or *ubi:RSdnM* or *hsp70l:RSN*. Confocal slices at the level of the AV canal were analysed. Non-trabecular tissue in these images was masked manually and then the total number of fluorescent pixels was measured using the IDL program (Research Systems). All images were taken at the same dimensions. Ventricle area was determined by measuring the total pixels outlined in the ventricle region using ImageJ software.

Clonal analysis. Cardiomyocyte clones were genetically labelled by combining the *myl7:CreER* and *priZm(β-act2:Brainbow1.0L)⁴* transgenes and then treating with 4-hydroxytamoxifen (4-HT, Sigma). Specifically, zebrafish embryos with these transgenes were treated at 48 hpf, when the zebrafish heart consists of a single cardiomyocyte thick wall and is looped but has not initiated cardiac trabeculation,

with 10 μ M 4-HT or 0.1% ethanol (control) for 6 h at 28 °C and then washed with fresh egg water several times. The dose and length of incubation of 4-HT was titrated to create small distinct clones (one or two cells) before trabeculation (Extended Data Fig. 8). The total numbers of cardiomyocyte clones were counted from three-dimensional reconstructions of confocal slices from hearts containing *myl7:CreER* and *priZm* transgenes. Visualization and counting of clones solely within the trabeculae were analysed using the post-imaging processing procedure described in the 'Image processing and statistical analysis' section below.

Mosaic analysis by DNA injection. To create the *hsp70l:loxP-mCherry-STOP-loxP-dnSuH-P2A-Emerald* plasmid (abbreviated as *hsp70l:RSdnS*), the dominant negative Suppressor of Hairless¹⁵ (*dnSuH*) DNA construct was PCR amplified with flanking *AscI* and *SacII* restriction sites. After sequence verification, this *dnSuH* product was subcloned into the *hsp70l:loxP-mCherry-STOP-loxP-NICD-P2A-Emerald*¹⁴ construct (*hsp70l:RSN*), replacing the NICD sequence and generating the *hsp70l:loxP-mCherry-STOP-loxP-dnSuH-P2A-Emerald* (*hsp70l:RSdnS*) construct for subsequent injection studies (see below). To generate cardiomyocyte clones with constitutively activated or inhibited Notch signalling, the *I-SceI* enzyme was co-injected with either the *hsp70l:RSN* plasmid (25 pg) or the *hsp70l:RSdnS* (25 pg) plasmid into one-cell stage embryos containing *Tg(myl7:Cre; myl7:Cerulean)* or only *Tg(myl7:Cerulean)*. Embryos were then heat-shocked (as described above) at 60 hpf and imaged at 72 hpf. Cardiomyocytes containing either plasmid were detected by the co-expression of mCherry or Emerald and Cerulean. The location of Notch-altered Emerald⁺ cardiomyocytes clones in either the ventricular outer wall or the trabeculae was determined using the method described within the 'Image processing and statistical analysis' section below.

Erbb2 and BMP loss of function studies. The activity of Erbb2, a tyrosine kinase receptor, was inhibited using (1) homozygous *erbb2*^{st50} mutants¹⁸, (2) a splice morpholino targeting *erbb2* (*erbb2* MO)¹⁰, or (3) the tyrosine kinase inhibitor AG1478 (Calbiochem). (1) The *erbb2*^{st50} homozygous mutant embryos were identified by the previously characterized aberrant cardiac morphology¹⁸. (2) The *erbb2* MO was previously characterized and shown to be specific¹⁰. We injected 570 pg of the *erbb2* morpholino (*erbb2* MO) or a mismatched control morpholino (control MO) into one-cell stage embryos as previously described¹⁰. (3) AG1478 (5 μ M; Calbiochem) or 0.1% DMSO (control) was added to embryos as previously described¹⁰. After incubation, embryos were washed extensively with egg water for further analysis. The *erbb2* MO injections or AG1478 incubations phenocopied the *erbb2*^{st50} mutant phenotype (Fig. 3 and Extended Data Fig. 9).

To investigate the relationship between Notch signalling and *erbb2*, *erbb2* MO embryos or *erbb2*^{st50} mutant embryos were incubated with 100 μ M DAPT (Sigma) or 0.1% DMSO at 60 hpf as described above. In complementary experiments, *Tg(hsp70l: dnMAML-GFP)* embryos were injected with the *erbb2* MO and heat-shocked at 60 hpf as described above.

To inhibit BMP signalling, embryos were incubated with 30 μ M of Dorsomorphin (Sigma) as previously described⁴¹. The efficacy of Dorsomorphin was verified by examining its effect on the BMP reporter, *Tg(BRE:d2GFP)*.

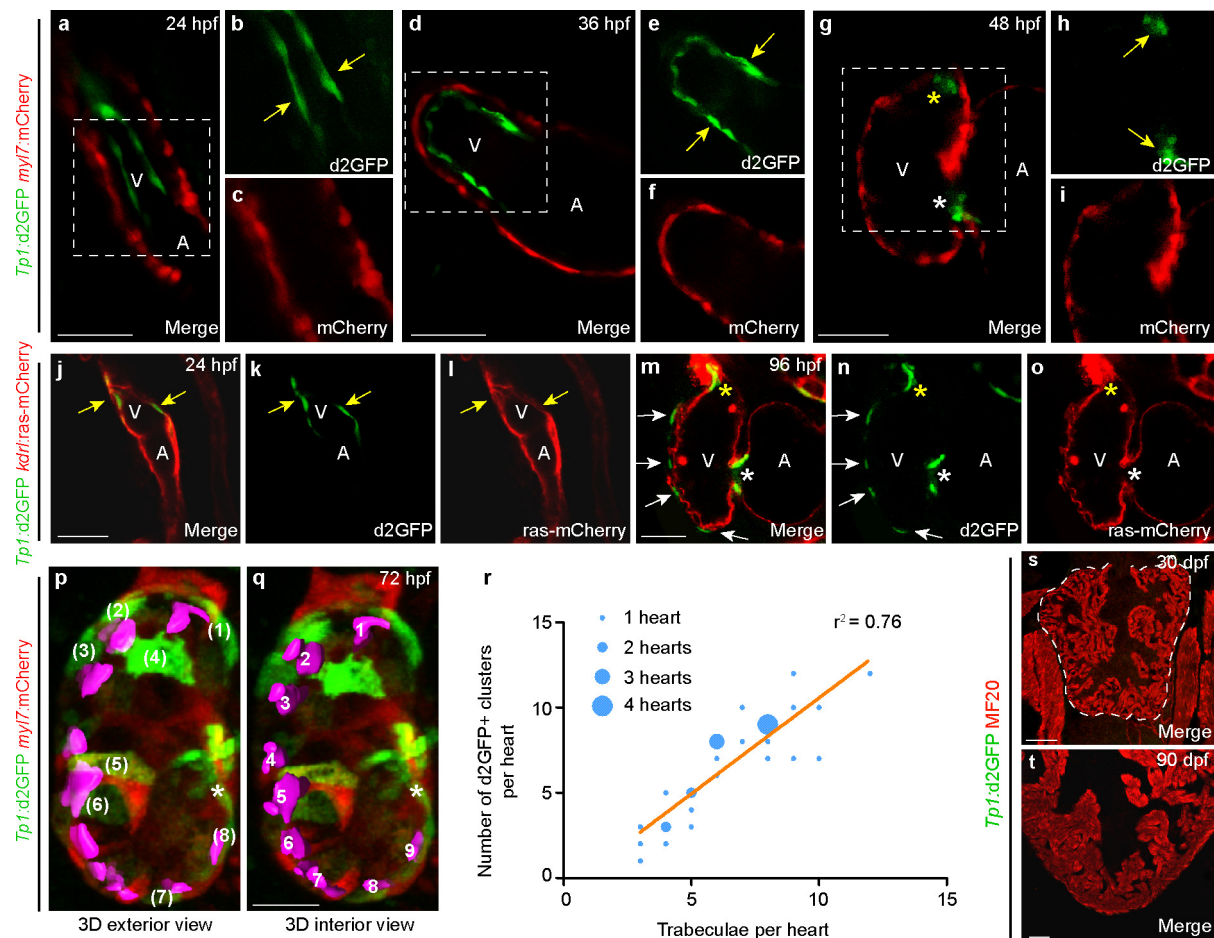
Cell transplantation studies. Blastomere transplantation was performed at the mid-blastula stage as previously described²³. Ten to twenty cells were removed from mid-blastula donor *Tg(myl7:Cerulean)* embryos and placed along the margin of either control MO or *erbb2* MO host *Tg(Tp1:d2GFP; myl7:H2A-mCherry)* embryos. Transplanted embryos in which donor cells contributed to the heart were imaged at 72 hpf. Image analysis of whether donor cells contributed to the ventricular outer wall or trabeculae and their location relative to *Tp1:d2GFP*⁺ cells was assessed in single confocal slices or in three-dimensional reconstructions using the post-imaging procedures described within the 'Image processing and statistical analysis' section below.

In situ hybridization expression analyses. Fluorescent *in situ* hybridization studies of *erbb2* and *jag2b* were performed as described in the ViewRNA *in situ* hybridization 1-Plex kit protocol (Affymetrix Panomics) with the following modifications. After the initial fixation, existing fluorescence was quenched as described for embryonic immunofluorescence above. Embryos were then subjected to protease digestion (protease QF, 1:100) at 40 °C for 30 min, followed by PBST washes and

re-fixation in 4% PFA at 25 °C for 20 min. After additional PBST washes, hybridization was performed with *erbb2* probes (Affymetrix Panomics VF1-16871, 1:50 dilution) or *jag2b* probes (Affymetrix Panomics VF1-18462, 1:50 dilution) at 40 °C overnight. Hybridized embryos were then washed in PBST and stepped through pre-labelling solutions. Embryos were then incubated with label probe-AP solution (1:1,000) for 30 min at 40 °C, washed again, transferred to a AP-enhancer solution and then transferred to fast red solution (one fast red substrate tablet in 5 ml naphthol buffer) for 30 min at 40 °C. After PBST washes, embryos were incubated in anti-MHC/MF20 antibody (mouse, Developmental Studies Hybridoma Bank, 1:100) or anti-GFP antibody (chicken, Aves Labs, 1:200) overnight at 4 °C. After PBST washes, embryos were incubated in anti-mouse IgG-Alexa 488 (goat, Life Technologies, 1:200) or anti-chicken IgG-Alexa 488 antibodies (goat, Life Technologies, 1:200) for 1 h at 25 °C. Finally, embryos were washed and mounted for imaging. PBST washing consisted of three washes for 15 min each at 25 °C. For *gfp* mRNA expression analysis, wholemount *in situ* hybridization studies were performed as previously described⁴⁰.

Image processing and statistical analysis. All images were obtained using a Nikon C2 confocal microscope and processed using Nikon NIS Elements software and ImageJ as previously described⁴⁰. Scale bars for all images represent 25 μ m. Measurements comparing the ventricular outer wall with the trabeculae were performed with post-image processing of confocal slices. Visualization of all cardiomyocytes and clones within the ventricle (comprising both the ventricular outer wall and the trabeculae) were made using three-dimensional reconstructions (Nikon NIS Elements) of confocal slices. However, to visualize only the trabeculae, cardiomyocytes within the ventricular outer wall in individual confocal slices were identified by their outer location and orientation and then masked manually. Three-dimensional reconstructions with these masked confocal slices then allowed the visualization and measurement of trabeculae alone. Measurements for the ventricular outer wall alone were calculated by subtracting the measurements of the trabeculae from the total ventricular cardiomyocytes. No statistical methods were used to predetermine sample size. Animals were assigned to experimental groups using simple randomization, without investigator blinding. Unpaired two-tailed Student's *t*-tests or Fisher's exact tests were used to determine statistical significance. *P* < 0.05 was considered to be statistically significant, as indicated by an asterisk. Error bars, s.e.m.

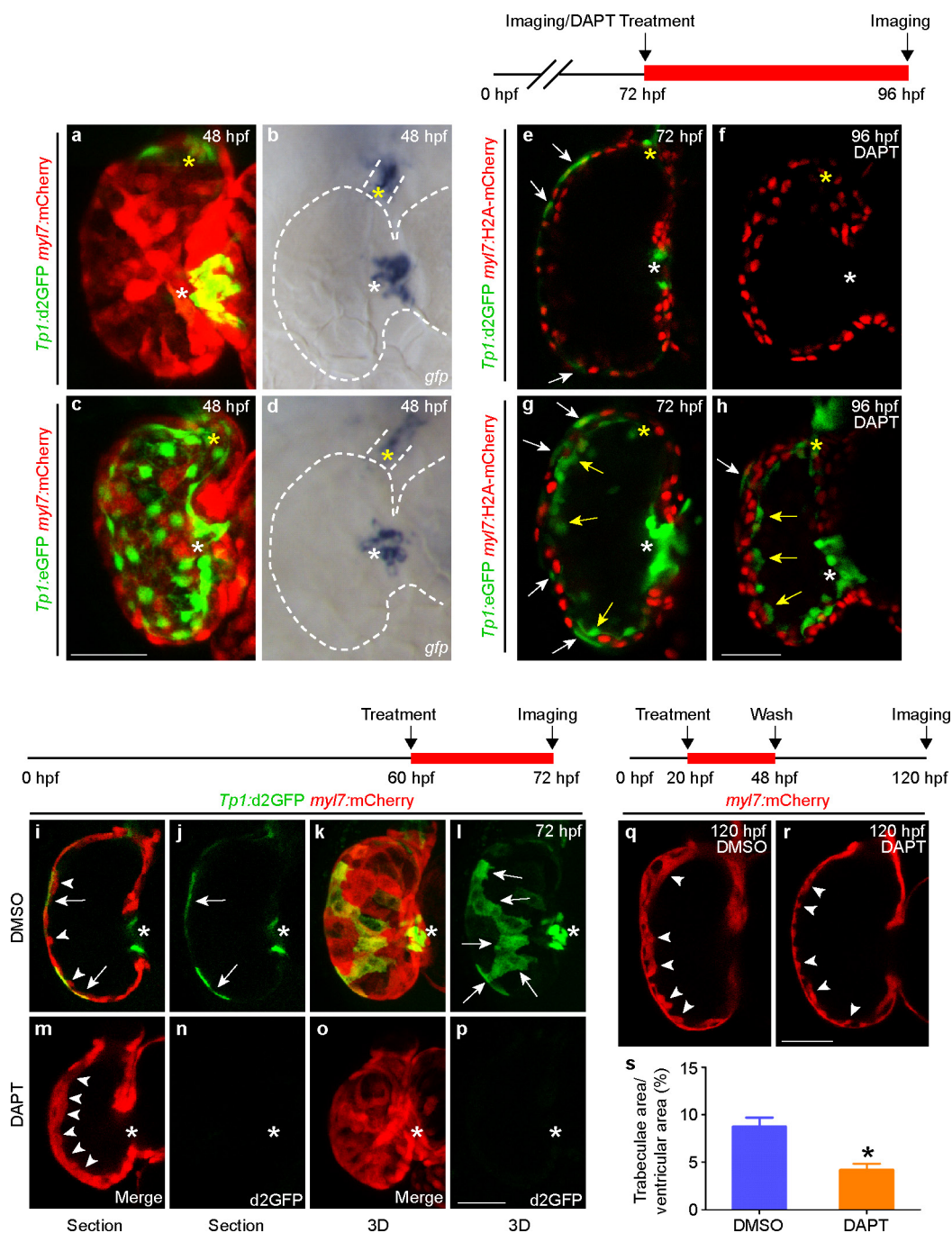
- Chi, N. C. *et al.* Foxn4 directly regulates *tbx2b* expression and atrioventricular canal formation. *Genes Dev.* **22**, 734–739 (2008).
- Palencia-Desai, S. *et al.* Vascular endothelial and endocardial progenitors differentiate as cardiomyocytes in the absence of *Etsrp/Etv2* function. *Development* **138**, 4721–4732 (2011).
- D'Amico, L., Scott, I. C., Jungblut, B. & Stainier, D. Y. A mutation in zebrafish *hmgcr1b* reveals a role for isoprenoids in vertebrate heart-tube formation. *Curr. Biol.* **17**, 252–259 (2007).
- Kikuchi, K. *et al.* Primary contribution to zebrafish heart regeneration by *gata4*⁺ cardiomyocytes. *Nature* **464**, 601–605 (2010).
- Kettleborough, R. N. *et al.* A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494–497 (2013).
- Huang, C. J., Tu, C. T., Hsiao, C. D., Hsieh, F. J. & Tsai, H. J. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**, 30–40 (2003).
- Thermes, V. *et al.* *I-SceI* meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev.* **118**, 91–98 (2002).
- Mosimann, C. *et al.* Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177 (2011).
- Kwan, K. M. *et al.* The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099 (2007).
- Zhou, Y. *et al.* Latent TGF- β binding protein 3 identifies a second heart field in zebrafish. *Nature* **474**, 645–648 (2011).
- Zhang, R. *et al.* In vivo cardiac reprogramming contributes to zebrafish heart regeneration. *Nature* **498**, 497–501 (2013).
- Yu, P. B. *et al.* Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nature Chem. Biol.* **4**, 33–41 (2008).



Extended Data Figure 1 | Notch signalling is dynamically activated in the endocardium and myocardium during heart development.

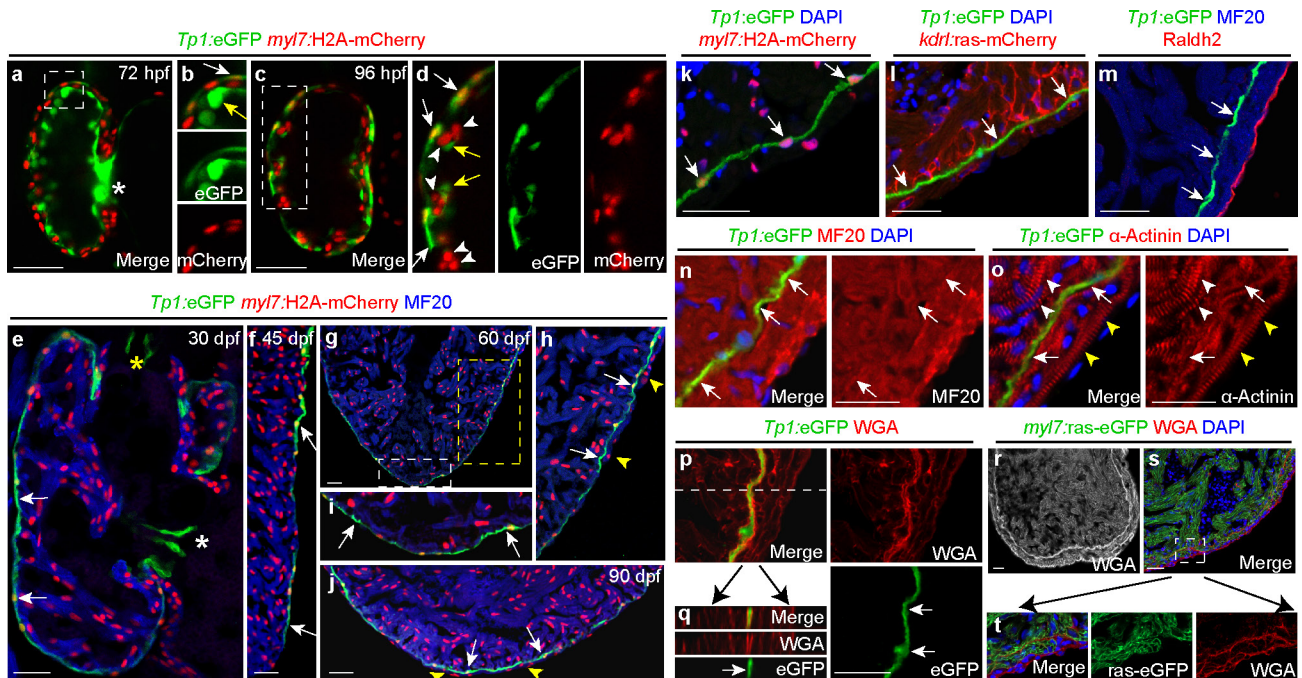
a–f, Confocal slices of *Tg(Tp1:d2GFP; myl7:mCherry)* hearts reveal that Notch signalling is in the ventricular endocardium (yellow arrows) but not in the myocardium at 24 hpf ($n = 11$) and 36 hpf ($n = 8$), but (**g–i**) becomes restricted to the AV and OFT endocardium by 48 hpf ($n = 12$). **j–o**, *Tg(Tp1:d2GFP; kdr1:ras-mCherry)* confocal imaging confirms that *Tp1:d2GFP* is expressed in the ventricular endocardium at (**j–l**) 24 hpf ($n = 8$) but becomes localized to the AV or OFT endocardium as well as non-endocardial cells in the outer ventricular myocardial wall (white arrows) by (**m–o**) 96 hpf ($n = 10$). **p, q**, Three-dimensional confocal reconstructions of the (p) exterior and (q) interior regions of 72 hpf *Tg(Tp1:d2GFP; myl7:mCherry)* hearts reveal that Notch-activated

Tp1:d2GFP⁺ cells are present in cardiomyocyte clusters (green, numbers in parentheses) and excluded from nascent cardiac trabeculae (pseudo-colour magenta, numbers). **r**, Graph shows that the number of cardiac trabeculae (x axis) and *Tp1:d2GFP*⁺ cardiomyocyte clusters (y axis) are similar within the ventricle ($n = 30$) at 72 hpf. Size of dots indicates the number of embryos with a particular number of trabeculae and *Tp1:d2GFP*⁺ clusters. Line represents a linear regression fitted to the data. **s, t**, Myocardial anti-MHC/MF20 immunostaining of *Tg(Tp1:d2GFP)* hearts reveals a loss of myocardial *Tp1:d2GFP* Notch reporter signal at 30 and 90 dpf hearts ($n = 5$ hearts per stage). White arrows, likely *Tp1:d2GFP*⁺ cardiomyocytes; yellow arrows, *Tp1:d2GFP*⁺ endocardial cells; white and yellow asterisks, AV and OFT. Dashed line in **s** outlines ventricle. V, ventricle; A, atrium. Scale bar, 25 μ m.



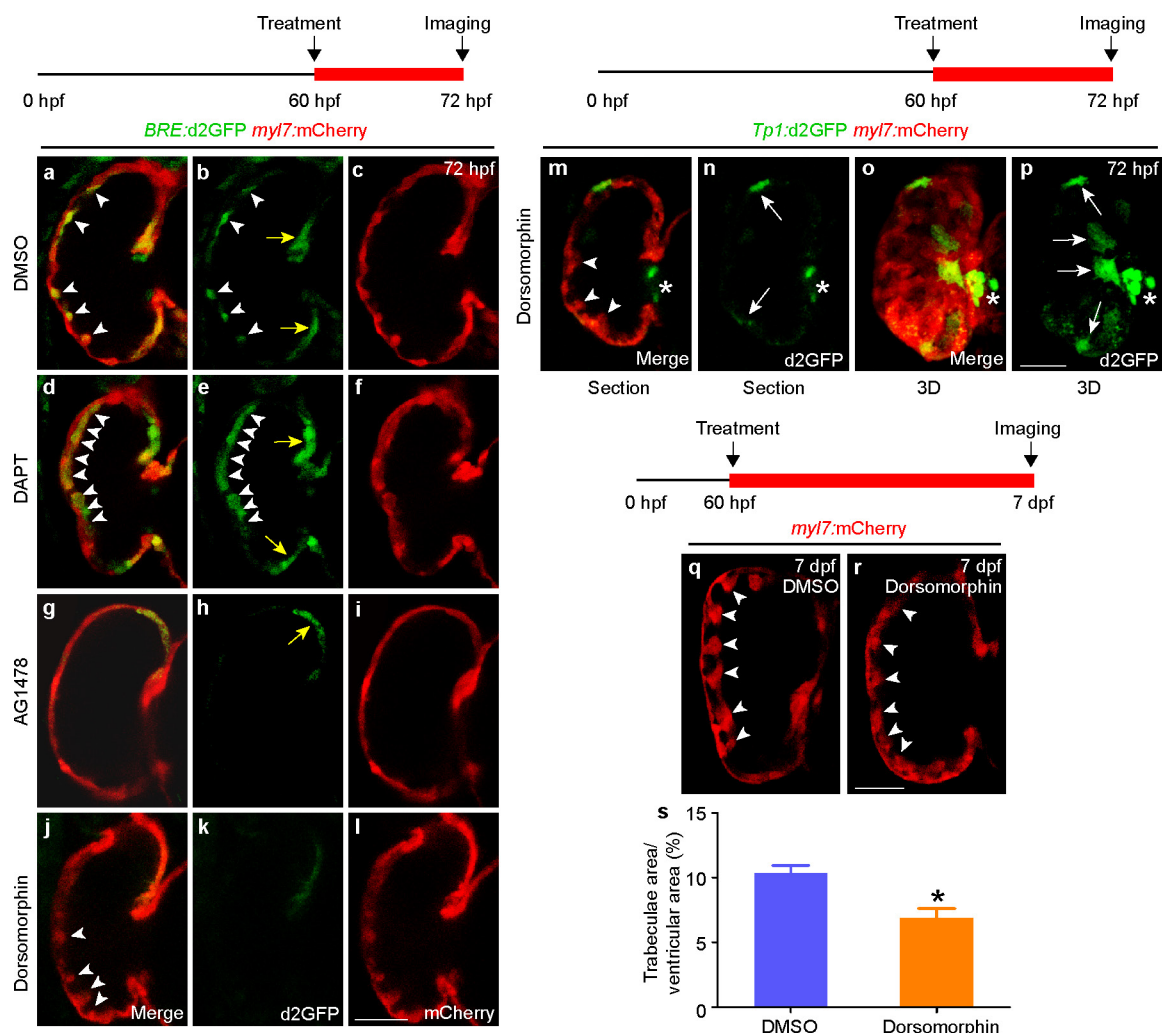
Extended Data Figure 2 | DAPT treatment validates that the Notch reporter *Tp1:d2GFP* monitors dynamic Notch signalling more closely than *Tp1:eGFP*, and reveals opposing roles of Notch signalling on trabeculation at different developmental stages. **a–d**, At 48 hpf, (a) *Tp1:d2GFP* expression is restricted to the AV and OFT endocardium ($n = 8/8$ embryos) whereas (c) *Tp1:eGFP* is expressed in the ventricular, AV and OFT endocardium ($n = 6/6$). However, *gfp* mRNA is primarily expressed in the AV and OFT regions in both (b) *Tg(Tp1:d2GFP; myl7:mCherry)* ($n = 10/10$) and (d) *Tg(Tp1:eGFP; myl7:mCherry)* embryos ($n = 5/5$), revealing that *Tp1:d2GFP* expression most closely matches Notch reporter activity. **e–h**, After 24 h DAPT treatments of (e, f) *Tg(Tp1:d2GFP; myl7:H2A-mCherry)* and (g, h) *Tg(Tp1:eGFP; myl7:H2A-mCherry)* embryos at 72 hpf, (f) *Tp1:d2GFP* is more diminished

throughout the heart at 96 hpf ($n = 8/10$) compared with (h) *Tp1:eGFP* ($n = 6/7$), confirming *Tp1:d2GFP* signal more faithfully recapitulates Notch signalling dynamics. **m–p**, *Tg(Tp1:d2GFP; myl7:mCherry)* hearts DAPT-treated from 60 to 72 hpf exhibit increased trabeculation (white arrowheads) and diminished *Tp1:d2GFP* Notch reporter activity ($n = 12/16$) than (i–l) DMSO-treated hearts ($n = 0/20$). However, (r) *Tg(myl7:mCherry)* hearts DAPT-treated from 20 to 48 hpf exhibit reduced trabeculae at 120 hpf ($n = 12/15$) than (q) DMSO-treated hearts ($n = 0/20$). **s**, Graph represents trabeculae/total ventricular area in embryos treated with DMSO or DAPT in q and r. White and yellow arrows, myocardial and endocardial Notch reporter activity; white arrowheads, trabeculae; white and yellow asterisks, AV and OFT. Scale bar, 25 μ m. Mean \pm s.e.m. * $P < 0.05$ by Student's *t*-test.



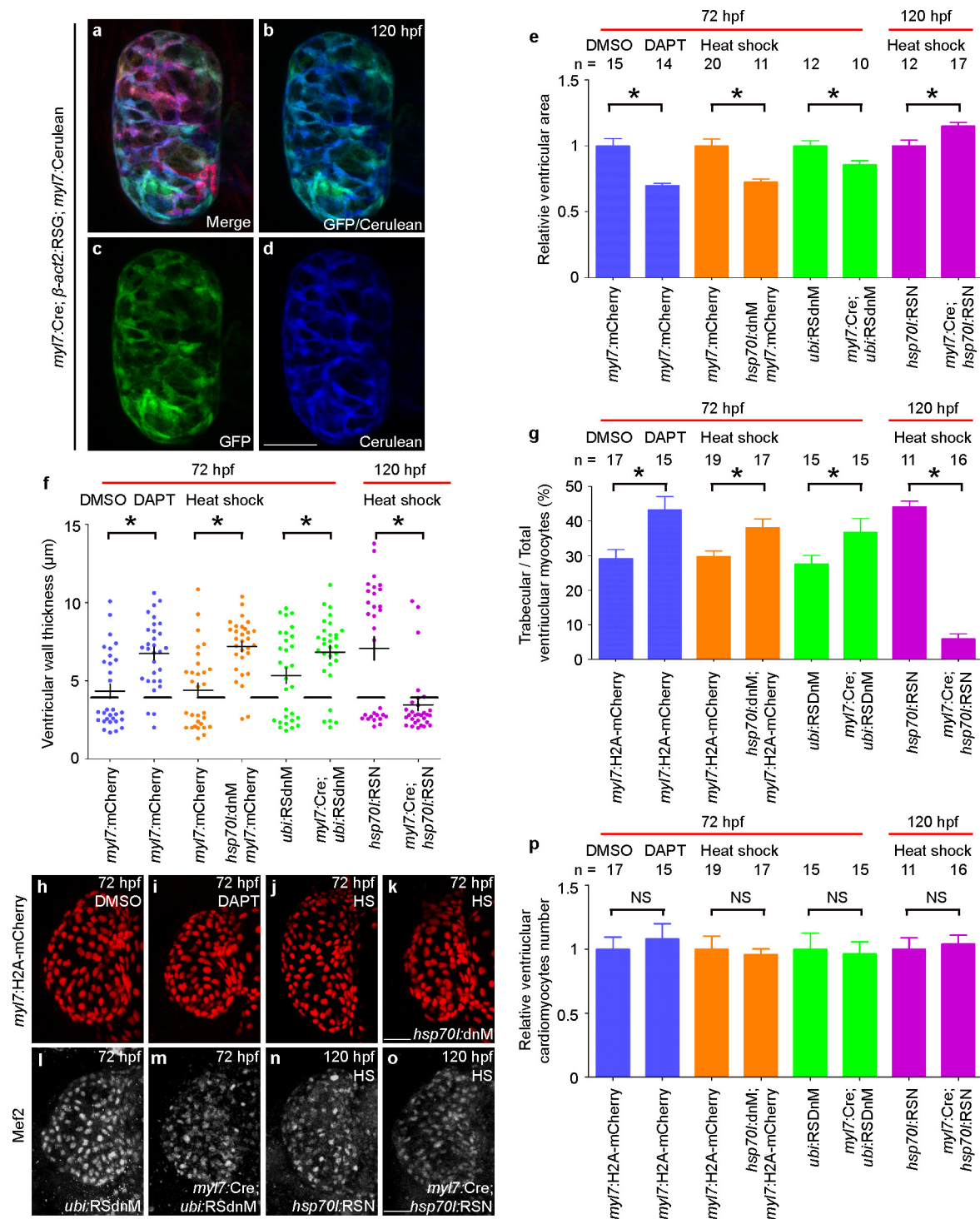
Extended Data Figure 3 | *Tpl1*:eGFP labels the ventricular outer wall during early cardiac development, which becomes the distinctive ventricular primordial myocardium in adults. Using the *Tpl1*:eGFP Notch reporter, which exhibits greater fluorescence perdurance than *Tpl1*:d2GFP, we performed limited fate mapping of Notch activated cardiac cells during ventricular morphogenesis. **a, b**, *Tpl1*:eGFP is expressed not only in ventricular cardiomyocytes (red nuclei, white arrows) at 72 hpf but also throughout the ventricular endocardium because of eGFP perdurance (yellow arrows) ($n = 12$). **c, d**, Although diminishing in the ventricular endocardium (yellow arrows) at 96 hpf ($n = 14$), *Tpl1*:eGFP expands in the outer ventricular myocardial wall (white arrows), yet is notably absent from myocardial trabeculae (white arrowheads). **e, f**, By 30 and 45 dpf ($n = 6$, $n = 5$), *Tpl1*:eGFP remains in the peripheral ventricular (primordial) myocardial layer, which is one cardiomyocyte thick (*myl7*:H2A-mCherry⁺/red and MF20⁺/blue), but is reduced in the ventricular but not the AV or OFT endocardium. **g–i**, At 60 dpf ($n = 5$), **(h)** new cardiomyocytes (cortical layer, yellow arrowheads) form over the *Tpl1*:eGFP⁺ primordial myocardium (white arrows) at the ventricular myocardial base (yellow box in **g**) and extend towards the apex where **(i)** *Tpl1*:eGFP⁺ cardiomyocytes (white arrows) still remain the outer most layer of the ventricular myocardium (white box in **g**). **j**, However, by 90 dpf ($n = 5$), this new cortical myocardial layer (yellow arrowheads)

spreads over the apical *Tpl1*:eGFP⁺ ventricular primordial myocardium (white arrows). **k–m**, In adult hearts (90 dpf), *Tpl1*:eGFP is primarily found in the (**k**, $n = 5$) *myl7*:H2A-mCherry⁺ primordial myocardium but not in the (**l**, $n = 5$) endocardium marked by *kdr*:ras-mCherry, nor (**m**, $n = 3$) epicardium marked by Raldh2 localization. **n–t**, Adult hearts (6 months) were further examined to assess the cellular attributes of the primordial layer. **n**, Anti-MHC/MF20 immunostaining confirms that *Tpl1*:eGFP⁺ cardiac cells are myocardial ($n = 5$). **o**, Anti- α -actinin immunostaining reveals that trabecular (white arrowheads) and cortical (yellow arrowheads) cardiomyocytes display organized sarcomeric structures but the *Tpl1*:eGFP⁺ primordial cardiomyocytes (arrows) do not ($n = 7$). **p–t**, Wheat germ agglutinin (WGA) staining shows that (**p, q**) the *Tpl1*:eGFP⁺ primordial myocardial layer is surrounded by extensive extracellular matrix ($n = 5$) and that (**r–t**) *Tg*(*myl7*:ras-eGFP) primordial cardiomyocytes display a thin cellular morphology compared with other ventricular cardiomyocytes ($n = 10$). **q**, An X–Z reconstruction of confocal stacks from *Tpl1*:eGFP and wheat germ agglutinin stainings at the dashed line shown in **p, b, d, h–i, t**, Magnifications of the boxed areas in **a, c, g, s**, respectively. White and yellow arrows, myocardial and endocardial *Tpl1*:eGFP; white and yellow arrowheads, trabeculae and cortical layer; white and yellow asterisks, AV and OFT. Scale bar, 25 μ m.



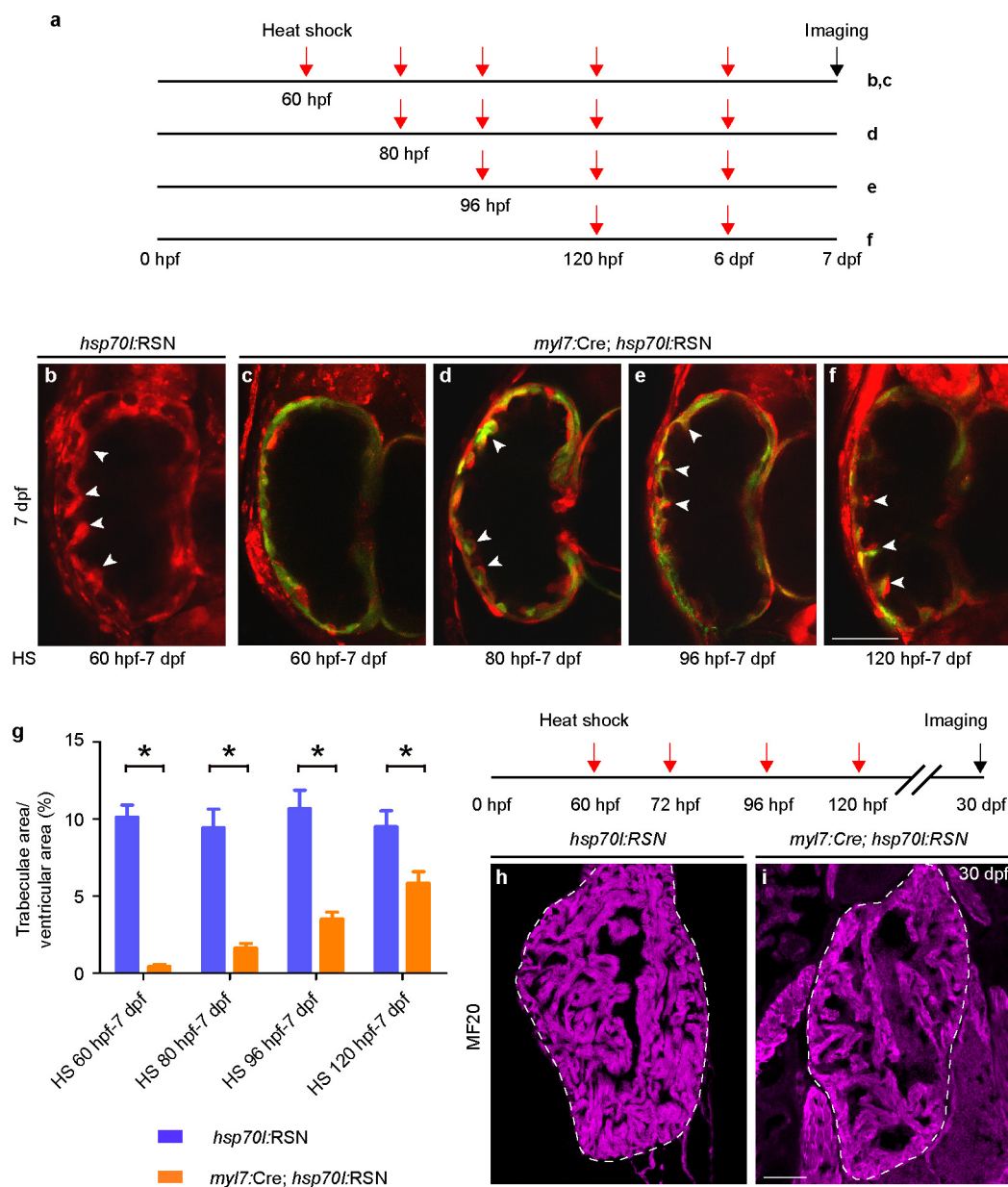
Extended Data Figure 4 | BMP signalling, which marks trabeculae, is required for expanding but not initiating trabeculae formation and has no effect on myocardial Notch activity. a–l, *Tg(BRE:d2GFP; myl7:mCherry)* hearts were treated with (a–c) DMSO, (d–f) DAPT, (g–i) AG1478, or (j–l) Dorsomorphin at 60 hpf and imaged at 72 hpf. a–c, DMSO-treated hearts express the *BRE:d2GFP* BMP reporter in trabeculae (arrowheads) and in the AV myocardium (yellow arrows, $n = 11/11$ embryos). d–f, DAPT-treated hearts exhibit increased trabeculation and *BRE:d2GFP* expression in these forming trabeculae (arrowheads, $n = 9/12$). g–i, AG1478-treated hearts fail to form trabeculae ($n = 9/10$) and only express the *BRE:d2GFP* BMP reporter in the AV myocardium (yellow arrow). j–l, Dorsomorphin-treated hearts form cardiac trabeculae (arrowheads) but fail to express the *BRE:d2GFP* BMP reporter in both cardiac trabeculae and the AV myocardium

($n = 10/12$). m–p, Treating *Tg(Tp1:d2GFP; myl7:mCherry)* embryos with Dorsomorphin from 60 to 72 hpf did not affect the initiation of trabeculae (arrowheads) nor the activation of myocardial Notch signalling (white arrows, $n = 13/16$) compared with treating with DMSO (see Extended Data Fig. 2i–l). q, r, Although *Tg(myl7:mCherry)* hearts treated with (q) DMSO or (r) Dorsomorphin from 60 hpf to 7 dpf form similar numbers of trabeculae (arrowheads), Dorsomorphin-treated hearts display trabeculae that are stunted/reduced in size ($n = 12/15$) compared with DMSO-treated control hearts ($n = 0/15$). s, Graph reveals a significant reduction in the trabecular/ventricular area ratio in Dorsomorphin-treated fish compared with DMSO-treated controls. Arrowheads, trabeculae; yellow arrows, AV myocardium; white arrows, *Tp1:d2GFP*⁺ myocardium. White asterisks, AV. Mean \pm s.e.m. * $P < 0.05$ by Student's *t*-test. Scale bar, 25 μ m.



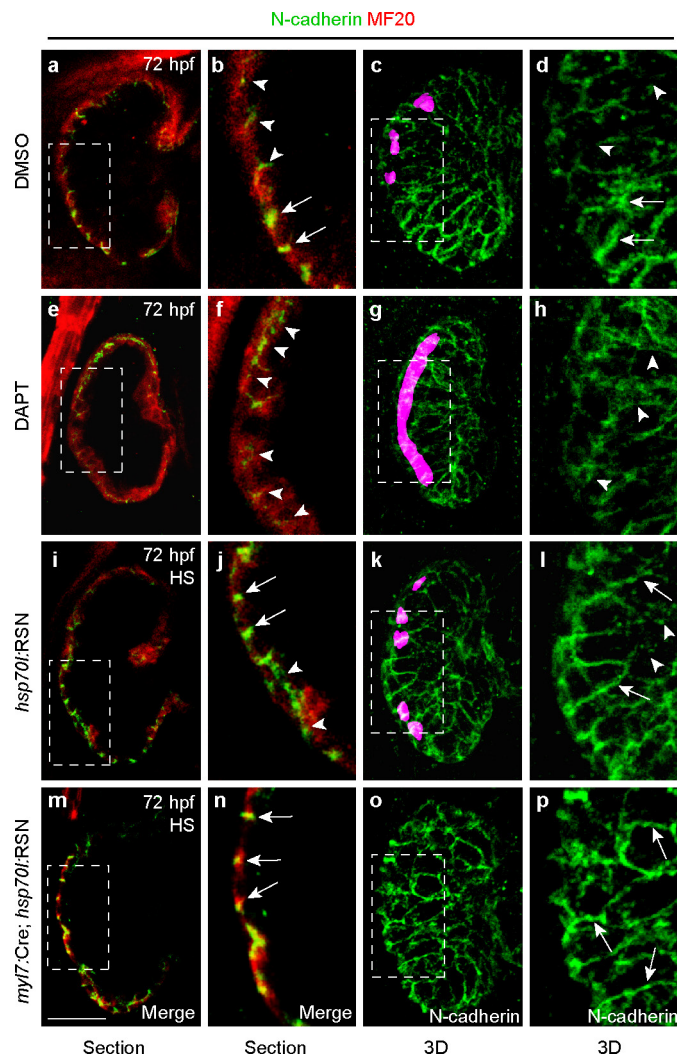
Extended Data Figure 5 | Altering myocardial Notch signalling affects ventricular size and wall thickness but not total number of ventricular cardiomyocytes. **a–d**, The *Tg(myl7:Cre)* transgenic line used to specifically perturb Notch signalling in the myocardium was validated by confirming that Cre expression is restricted to the myocardium. Activity of *myl7:Cre*, as visualized by (c) GFP expression from the switch line, β -act2:RSG, exclusively overlaps with (b, d) *myl7:Cre*:Cerulean expression at 120 hpf ($n = 10$ embryos). Quantitative analyses of (e) ventricular size and (f) wall thickness performed on confocal images from Fig. 2a–h reveal that myocardial Notch signalling restricts ventricular size while promoting ventricular wall thickness. **e**, Ventricular size measurements were normalized to respective controls for each condition. **f**, Individual measurements (dots) of myocardial thickness were taken across the outer curvature of the ventricle ($n = 30$ measurements, 6 measurements were

taken per embryo, 5 embryos per condition). Dashed line represents the ventricular wall thickness that distinguishes trabeculated myocardial thickness from ventricular outer wall myocardial thickness in control hearts. Crosses denote mean and s.e.m. **g–p**, Quantitative analysis of (g) trabecular cardiomyocytes and (p) total ventricular cardiomyocytes was calculated by counting myocardial nuclei labelled with *myl7:H2A-mCherry* or anti-Mef2 immunostaining using embryos from Fig. 2i–p for **g**, or from three-dimensional reconstructions in **h–o** for **p**. In **g**, the number of trabecular/total ventricular cardiomyocytes was used to calculate the percentage of trabecular cardiomyocytes for each condition. In **p**, total ventricular cardiomyocytes were normalized to respective controls for each condition. n , Number of embryos analysed per condition. Mean \pm s.e.m. * $P < 0.05$ by Student's *t*-test. NS, not significant. Scale bar, 25 μ m.

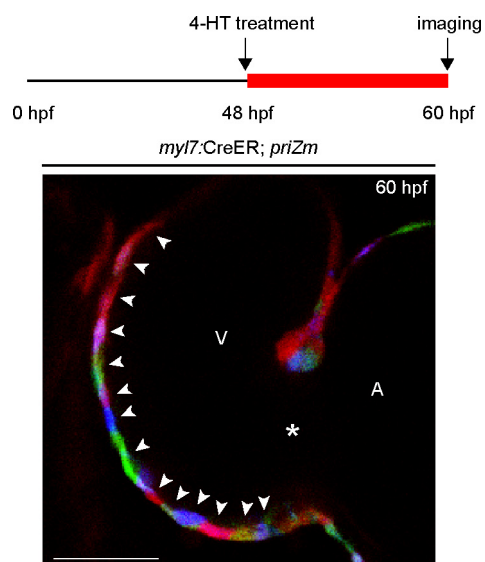


Extended Data Figure 6 | Myocardial Notch activation can inhibit the formation and expansion of cardiac trabeculae at various cardiac developmental stages. **a–g.** *Tg(myl7:Cre; hsp70l:RSN)* and *Tg(hsp70l:RSN)* (control) embryos were heat-shocked (HS) during various developmental time windows as indicated and imaged at 7 dpf to assess the effects of constitutive myocardial Notch signalling on cardiac trabeculae formation. **a.** Red arrows in schematic indicate the time points at which embryos in the corresponding panels were heat-shocked. **b.** Control *Tg(hsp70l:RSN)* embryos heat-shocked from 60 hpf to 7 dpf ubiquitously express mCherry but do not overexpress myocardial NICD. They form cardiac trabeculae (arrowheads) similar to wild-type embryos (control, $n = 14/15$). **c.** However, *Tg(myl7:Cre; hsp70l:RSN)* embryos heat-shocked from 60 hpf to 7 dpf overexpress NICD-P2A–Emerald throughout the myocardium and fail to form cardiac trabeculae ($n = 9/12$). Although

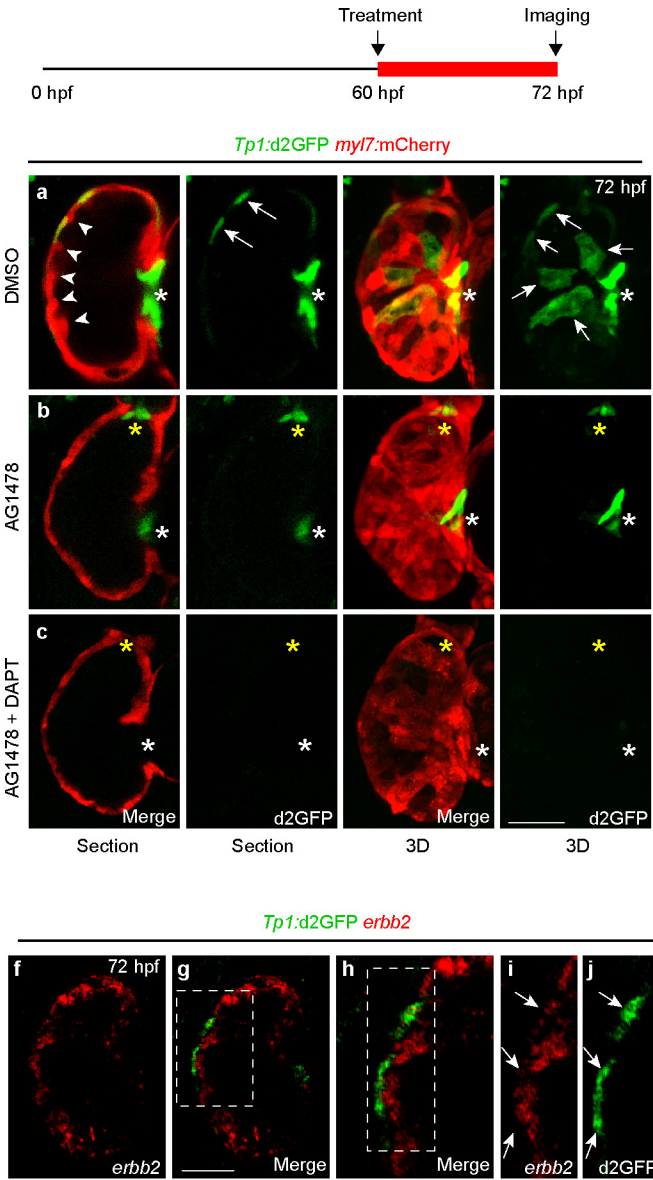
Tg(myl7:Cre; hsp70l:RSN) embryos heat-shocked at **(d)** 80 hpf, **(e)** 96 hpf, and **(f)** 120 hpf form trabeculae, these embryos exhibit stunted/smaller trabeculae after heat-shocking ($n = 9/10$, $10/14$, and $12/16$, respectively). **g.** Graph of trabeculae/total ventricular area of heat-shocked embryos from **b–f**, showing that myocardial Notch over-activation inhibits the progression of cardiac trabeculae formation. **h, i.** Although heat-shocking *Tg(myl7:Cre; hsp70l:RSN)* from 60 to 120 hpf initially inhibits trabeculae formation, **(i)** the ventricular myocardium (detected by anti-MHC/MF20 immunostaining, magenta) can still form trabeculae, albeit at reduced numbers ($n = 4/5$) by 30 dpf after stopping NICD overexpression compared with **(h)** heat-shocked *Tg(hsp70l:RSN)* hearts (control, $n = 0/8$). HS, heat-shock; white arrowheads, trabeculae. Scale bar, 25 μ m. Mean \pm s.e.m. * $P < 0.05$ by Student's t -test.



Extended Data Figure 7 | Notch signalling regulates cardiomyocyte cell junctions during cardiac trabeculae formation. **a–d**, In DMSO-treated (control) 72 hpf wild-type hearts, N-cadherin is localized at cell junctions of cardiomyocytes within the ventricular outer wall (arrows) but redistributes away from these cell–cell contacts in cardiomyocytes that extend into the lumen to form trabeculae (arrowheads) ($n = 12/12$). **e–h**, Notch inhibition by DAPT treatment promotes N-cadherin redistribution and results in increased trabeculation ($n = 8/11$). **m–p**, Conversely, myocardial Notch activation by heat shocking (HS) *Tg(myl7:Cre; hsp70i:RSN)* leads to diminished N-cadherin redistribution and reduced trabeculation ($n = 7/10$) compared with (**i–l**) heat-shocked *Tg(hsp70i:RSN)* control hearts ($n = 0/10$). Nascent cardiac trabeculae were pseudo-coloured magenta in **c, g** and **k**. **b, d, f, h, j, l, n, p**, Magnifications of boxed areas in **a, c, e, g, i, k, m, o**, respectively. Arrowheads, N-cadherin redistributed from cell–cell contacts; arrows, N-cadherin at cell–cell contacts within outer wall. Scale bar, 25 μm .

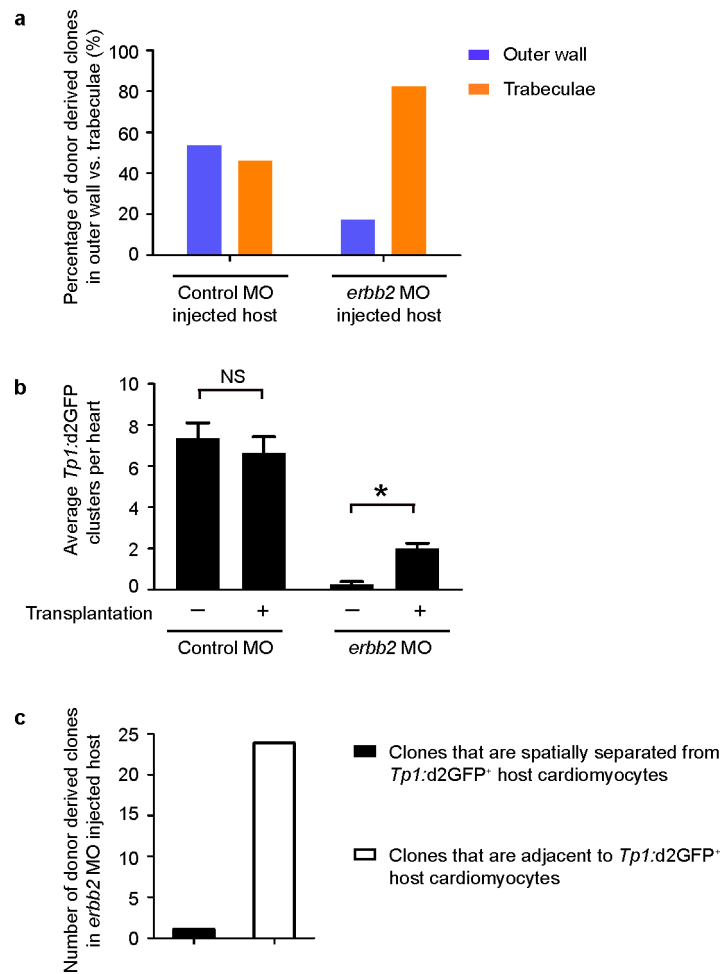


Extended Data Figure 8 | Tamoxifen treatment of *Tg(myl7:CreER; priZm)* embryos at 48 hpf labels adjacent individual cardiomyocytes with combinations of distinct fluorescent colours. *Tg(myl7:CreER; priZm)* embryos were treated with 4-HT at 48 hpf and confocal imaged at 60 hpf before the initiation of cardiomyocytes forming trabeculae. Individual cardiomyocytes (arrowheads) are labelled with distinct combinations of fluorescent proteins allowing for tracking of specific cardiomyocyte clones ($n = 6$). White arrowheads, cardiomyocytes; V, ventricle; A, atrium; white asterisk, AV. Scale bar, 25 μm .



Extended Data Figure 9 | Notch and ErbB2 signalling pathways form a feedback loop during cardiac trabeculation. a, b, Compared with (a) DMSO-treated *Tg(Tp1:2dGFP; myl7:mCherry)* (controls) embryos, (b) inhibiting ErbB2 function with AG1478 from 60 to 72 hpf blocks trabeculation and myocardial Notch signalling ($n = 14/17$), confirming *erbB2* MO and mutant phenotypes. **c,** However, Notch inhibition using DAPT cannot reverse the AG1478/ErbB2 inhibition effect on trabeculae formation ($n = 11/12$). **d, e,** Consistent with these results, (d) control MO-injected *Tg(hsp70l:dnM; myl7:mCherry)* embryos expressing heat-shock induced dnMAML from 60 to 72 hpf display increased trabeculation (arrowheads, $n = 9/11$); (e) however, *erbB2* MO-injected embryos expressing heat-shock induced dnMAML fail to display trabeculae ($n = 9/12$) as similarly observed in *erbB2* MO-injected embryos alone (Fig. 3). **f–j,** The *erbB2* fluorescent *in situ* hybridization and GFP

co-immunostaining performed on 72 hpf *Tg(Tp1:d2GFP)* hearts reveal that *erbb2* is expressed in an intermittent pattern across the ventricular wall and is specifically diminished in *Tp1:d2GFP⁺* cells (arrows) ($n = 6/6$). **l, p**, Heat-shocked (HS) *Tg(myl7:Cre; hsp70l:RSN)* hearts, which exhibit constitutively activated myocardial Notch signalling (NICD) from 60 to 120 hpf, minimally express *erbb2* in the myocardium ($n = 8/11$) compared with **(k, o)** heat-shocked *Tg(hsp70l:RSN)* control hearts ($n = 0/20$) at 120 hpf. Compared with **(m, q)** DMSO-treated control hearts ($n = 0/10$), **(n, r)** Notch-inhibited hearts by DAPT treatment from 60 to 72 hpf exhibit increased myocardial *erbb2* expression as well as more trabeculae at 72 hpf ($n = 8/10$), supporting the idea that Notch signalling inhibits *erbb2* expression. **h, i–j**, Magnifications of boxed areas in **g, h**, respectively. Arrowheads, trabeculae; arrows, *Tp1:d2GFP⁺* cardiomyocytes; white and yellow asterisks, AV and OFT. Scale bar, 25 μm .



Extended Data Figure 10 | Transplanted wild-type cardiomyocytes non-cell-autonomously activate Notch signalling in *erb2* morphant host cardiomyocytes. **a**, On the basis of mosaic embryo studies from Fig. 3f–i, wild-type donor cardiomyocytes contribute equally to the outer ventricular wall (14/26 clones) or the trabeculae (12/26 clones) when transplanted into control MO host embryos ($n = 12$ embryos). However, when wild-type donor cells are transplanted into *erb2* MO host embryos ($n = 10$ embryos), they contribute more to the trabecular layer (19/23 clones) than to the ventricular outer wall (4/23 clones, $P < 0.05$ by Fisher's exact test). **b**, On the basis of mosaic embryo studies from Fig. 3f–i,

transplanting wild-type donor cells increases the number of *erb2* MO host cardiomyocytes expressing *Tp1:d2GFP* ($n = 10$ embryos) compared with non-transplanted *erb2* MO embryos ($n = 16$ embryos), but had no effect on the number of control MO host cells expressing *Tp1:d2GFP* ($n = 12$ embryos) compared with non-transplanted controls ($n = 11$ embryos). **c**, Quantitative data for Fig. 3f–i reveal that transplanted wild-type donor cardiomyocytes are primarily adjacent to host *Tp1:d2GFP*⁺ cardiomyocytes in *erb2* MO hearts ($n = 10$ embryos). Mean \pm s.e.m. $*P < 0.05$ by Student's *t*-test. NS, not significant.

Widespread transmission of independent cancer lineages within multiple bivalve species

Michael J. Metzger^{1,2}, Antonio Villalba^{3,4}, María J. Carballal³, David Iglesias³, James Sherry⁵, Carol Reinisch⁵, Annette F. Muttray^{6,7}, Susan A. Baldwin⁶ & Stephen P. Goff^{1,2,8}

Most cancers arise from oncogenic changes in the genomes of somatic cells, and while the cells may migrate by metastasis, they remain within that single individual. Natural transmission of cancer cells from one individual to another has been observed in two distinct cases in mammals (Tasmanian devils¹ and dogs^{2,3}), but these are generally considered to be rare exceptions in nature. The discovery of transmissible cancer in soft-shell clams (*Mya arenaria*)⁴ suggested that this phenomenon might be more widespread. Here we analyse disseminated neoplasia in mussels (*Mytilus trossulus*), cockles (*Cerastoderma edule*), and golden carpet shell clams (*Polititapes aureus*) and find that neoplasias in all three species are attributable to independent transmissible cancer lineages. In mussels and cockles, the cancer lineages are derived from their respective host species; however, unexpectedly, cancer cells in *P. aureus* are all derived from *Venerupis corrugata*, a different species living in the same geographical area. No cases of disseminated neoplasia have thus far been found in *V. corrugata* from the same region. These findings show that transmission of cancer cells in the marine environment is common in multiple species, that it has originated many times, and that while most transmissible cancers are found spreading within the species of origin, cross-species transmission of cancer cells can occur.

Disseminated neoplasia, or haemic neoplasia, a leukaemia-like disease, occurs with high prevalence in multiple bivalve species^{5,6}. Here we investigate the possibility that cancers in three species could be attributed to transmissible cancer cells, and whether these cancers are restricted to the species of origin or can undergo cross-species transmission.

Mussels (*M. trossulus*; Fig. 1a) are subject to disseminated neoplasia in the Pacific Northwest Coast^{7,8}, and evidence of common polymorphisms in neoplasias suggested that these might represent a transmissible cancer⁹. Twenty-eight mussels (*M. trossulus*) collected from West Vancouver were screened for neoplasia by drawing haemolymph and analysing haemocytes for the rounded, non-adherent morphology of neoplastic cells. Two were identified with high levels of neoplastic cells. We sequenced part of the mitochondrial cytochrome *c* oxidase I (*mtCOI*) gene in host tissue and neoplastic haemocytes from the two diseased animals and four normal animals to test whether the neoplastic cells were derived from the host individuals or exhibited a distinct genotype, the hallmark of transmissible neoplasia. While solid tissue and haemocyte genotypes within each normal animal were always identical, solid tissue and neoplastic haemocyte genotypes of the diseased animals were discordant (Fig. 1b and Extended Data Fig. 1). Moreover, the same single nucleotide polymorphisms (SNPs) were found in neoplastic cells of the two diseased individuals, indicating that the cancer cells were not of host origin and suggesting that they arose from a single clonal origin. *EF1 α* gene sequences also revealed that the genotypes of the host cells and neoplastic haemocytes were discordant, and that the

genotypes of the neoplastic cells of the two different animals were again identical to each other (Fig. 1c).

To determine whether this transmissible cell line was widespread in the *M. trossulus* population, 250 were collected from Vancouver Island, and seven potentially diseased individuals were analysed. In one neoplastic sample (MW81), the haemocyte and tissue genotypes did not match, and the haemocytes contained the same *mtCOI* allele and the same *EF1 α* major and minor alleles found in the other samples (Fig. 1b, c and Extended Data Fig. 1). These genotypes strongly indicate the existence of a *M. trossulus*-derived transmissible cancer lineage circulating in the wild population.

High prevalence of disseminated neoplasia has been observed in two species of bivalves on the Galician Coast: cockles (*C. edule*)^{10,11} and golden carpet shell clams (*P. aureus*, previously named *Venerupis aurea*)¹². The disease in cockles (*C. edule*) exhibits one of two distinct morphologies, termed types A and B^{13,14}. We collected about 150 cockles (*C. edule*; Fig. 2a), and examined the genotypes of solid tissue and haemocytes of six normal individuals and six with high (>75%) or moderate (15–75%) amounts of neoplastic cells in the haemolymph. Nine polymorphic microsatellite loci¹⁵ were amplified from normal animals, and allele sizes from tissue and haemocytes of each normal animal all matched, but allele sizes in haemocytes and tissue of diseased animals were discordant (Fig. 2b, c). In a phylogenetic tree based on microsatellite alleles^{16,17} the neoplastic haemocyte genotypes did not group with the host tissue genotypes, consistent with transmissible cancer, and instead clustered into two distinct branches, suggesting two independent cancer lineages (Fig. 2d). We sequenced the *mtCOI* gene and identified several SNPs that were present only in lineage 2, and not in any of the normal animals. No unique SNPs were identified in the *mtCOI* region sequenced in lineage 1.

The microsatellite alleles and *mtCOI* SNPs suggest two independent cancer lineages, but these data are also consistent with two subgroups that have diverged from a single transmissible cancer lineage. To investigate this question, we sequenced an approximately 3-kilobase intron-spanning region in *EF1 α* from six normal individuals and two diseased individuals from each lineage (Fig. 2e). Both neoplastic haemocyte alleles were different from the tissue alleles in all diseased individuals, and the two alleles of the neoplasm in each diseased individual were nearly identical to the two alleles of the other individual in the lineage. However, both alleles in lineage 1 cells were different from those in lineage 2. Moreover, the neoplastic alleles were more closely related to some normal alleles than they were to the alleles of the alternative lineage. These data strongly suggest an independent origin of these two lineages.

Histological and morphometric examination showed that all three samples in lineage 1 were type A (characterized by a looser arrangement of neoplastic cells in the connective tissue, with pleomorphic nuclei) and all three samples in lineage 2 were type B (tighter arrangement,

¹Department of Biochemistry and Molecular Biophysics, Columbia University, New York 10032, USA. ²Howard Hughes Medical Institute, New York, New York 10032, USA. ³Centro de Investigacións Mariñas, Consellería do Mar, Xunta de Galicia, Vilanova de Arousa 36620, Spain. ⁴Department of Life Sciences, University of Alcalá, Alcalá de Henares 28871, Spain. ⁵Environment Canada, Water Science & Technology Directorate, Burlington, Ontario L7R 4A6, Canada. ⁶Chemical and Biological Engineering, University of British Columbia, Vancouver V6T 1Z3, Canada. ⁷SLR Consulting Canada Ltd., Vancouver V6J 1V4, Canada. ⁸Department of Microbiology and Immunology, Columbia University, New York, New York 10032, USA.

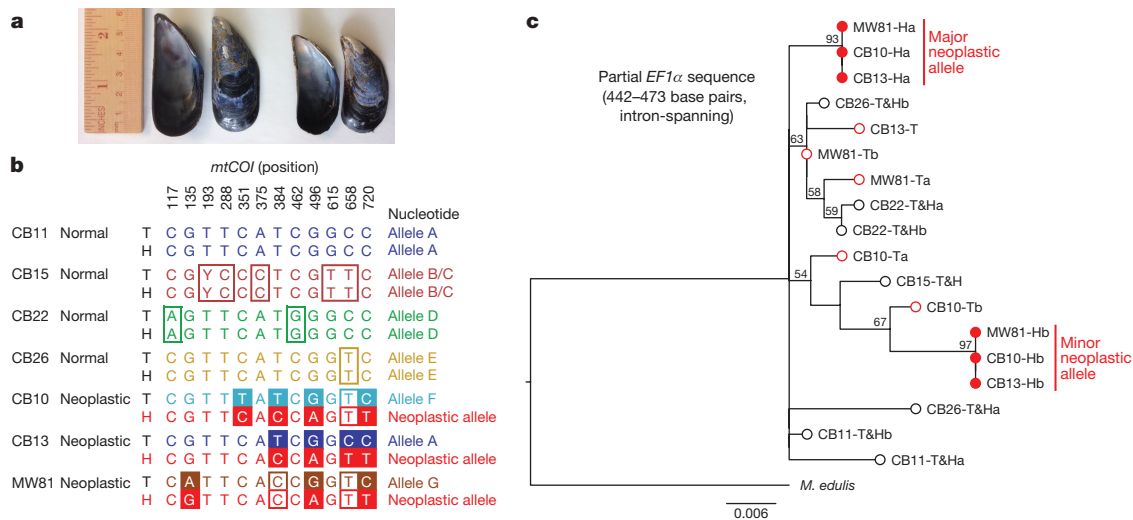


Figure 1 | Analysis of tissue and haemocyte genotypes of normal mussels (*M. trossulus*) and mussels with disseminated neoplasia using mitochondrial and nuclear DNA markers. a, Representative *M. trossulus*, with ruler for scale. b, Sequencing of *mtCOI* (bases 99–759) was conducted for both host solid tissue (T) and haemocytes (H) of normal ($n = 4$) and neoplastic mussels ($n = 3$). Open boxes mark SNPs that differ from allele A. Filled boxes mark discordance between neoplastic haemocyte and tissue genotypes. Sequences are numbered using AY823625.1. c, Maximum likelihood phylogenetic tree based on sequences of an intron-containing

region of *EF1α*, with bootstrap values over 50 shown and a scale bar showing genetic distance. *M. edulis* (614 base pairs (bp), EU684203.1) was used as an outgroup. In normal animals, the sequences of tissue and haemocyte DNA were identical and are presented together (T&H; black circles). For neoplastic animals, the tissue (T; open red circles) and haemocyte (H; filled red circles) alleles differed. Letters 'a' and 'b' denote multiple alleles in heterozygous individuals, and major and minor alleles from neoplastic alleles are marked. See Extended Data Fig. 1 for further details.

with smaller cells than type A and rounded, smaller nuclei; Fig. 2f–i and Extended Data Table 1). Altogether, these results argue that two distinct lineages of transmissible cancer, with distinct morphologies and genotypes, arose independently in cockles and are circulating in this species.

We also examined golden carpet shell clams (*P. aureus*; Fig. 3a), which are present in the same habitat as several other bivalves, including the closely related pullet shell clam (*V. corrugata*; Fig. 3b). Of 74 *P. aureus* individuals tested, 9 had high levels of disease and 22 had low-to-medium disease. We sequenced regions of *mtCOI*, the ribosomal DNA internal transcribed spacer (rDNA ITS), and *EF1α* from tissue and haemocyte DNA of six highly diseased and six normal *P. aureus* individuals, and again found that the genotypes of neoplastic cells were nearly identical and did not match those of their hosts (Fig. 3c–e). In contrast to the transmissible cancers in other species, however, the sequences of the neoplastic cells were highly dissimilar from normal host sequence (only 78.4–78.5% identical in the *mtCOI* locus and 89.3–93.2% identical at nuclear loci, ignoring insertions and deletions). The neoplastic genotypes were instead near-perfect matches to the sequences of *V. corrugata* (98.6–7% and 99.3–100% identical, respectively).

To confirm that detection of the *V. corrugata* sequence reflected the neoplastic cells observed morphologically, we used species-specific *EF1α* quantitative PCR (qPCR) to quantify the fraction of cells that were derived from each species in the tissue and haemocyte samples (Fig. 3f). As expected, no *V. corrugata* sequences were detected in normal *P. aureus* animals, but high frequencies of *V. corrugata* DNA were detected in all six individuals diagnosed with high levels of neoplasia and lower amounts in most individuals diagnosed with medium and low levels. The small number of individuals diagnosed with low and medium disease in which *V. corrugata* DNA was not detected could have primary host-derived neoplasia, a different transmissible cancer lineage, too low an abundance of tumour DNA to detect, or even a small amount of normal cells with an unusually rounded morphology. In diseased animals, the highest levels of *V. corrugata* sequence were detected in haemocytes, and lower levels were present in tissues, probably because of infiltration of cancer cells via the circulation. We conclude that the cancer spreading in the *P. aureus* population resulted

from cross-species transmission of cancer cells of *V. corrugata* origin. It is noteworthy that no cases of disseminated neoplasia have thus far been found in *V. corrugata* from the same region, despite analysis of hundreds of clams, suggesting that the *V. corrugata* cancer cells do not initiate disease in the species of origin and were only observed to colonize *P. aureus* animals.

We previously identified an LTR-retrotransposon, *Steamer*, which was amplified in the transmissible neoplasia lineage in the soft-shell clam (*M. arenaria*)¹⁸. Using degenerate primers, we identified at least one *Steamer*-like element (SLE) in each of the three species studied here. In each case, copy numbers were variable among individuals, but no massive amplification of these particular retrotransposons was observed in any of the transmissible cancers assayed (Extended Data Fig. 2). This suggests that *Steamer*-like retrotransposon amplification is not essential in development of transmissible clones.

Our results indicate that transmission of contagious cancer cells is a widespread phenomenon in the marine environment, with multiple independent lineages developing in multiple species in four bivalve families. Along with the recent identification of a second independent lineage of transmissible cancer in Tasmanian devils¹⁹, these findings confirm that, under suitable conditions, the development of transmissible cancer can occur multiple times, and suggests that some species may be more susceptible to development of transmissible neoplasia than others.

Spontaneous haemic neoplasia must occur at some frequency for transmissible lineages to arise, but cases of transmissible cancer appear to outnumber spontaneous disease, at least in the species investigated so far (that is, all neoplastic cases tested in cockles, golden carpet shell clams, and three of nine samples in mussels, in addition to all samples in the previous report of soft-shell clams⁴).

Transmissible cancers appear largely restricted to the species of origin in nature. While canine transmissible venereal tumour has been experimentally transplanted to coyotes²⁰, jackals²¹, and foxes^{22–24}, no examples of natural transmission beyond dogs have been reported²⁵. Disseminated neoplasias of molluscs have only been transplanted to members of the same species (soft-shell clams, mussels, and others^{5,6}). Attempts to transfer *M. arenaria* neoplasia through water exposure to both *M. arenaria* and *M. trossulus* only resulted in transfer

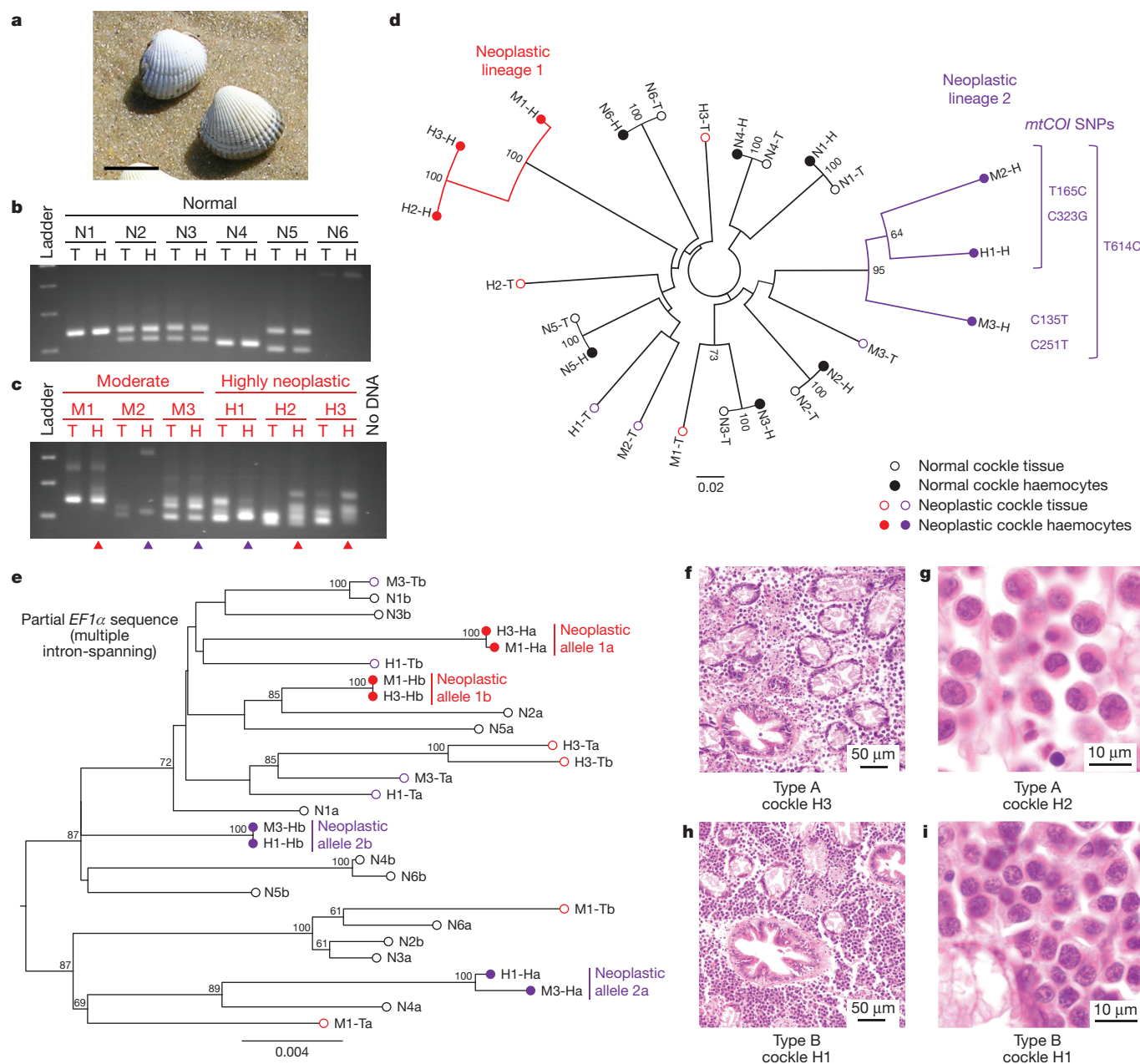


Figure 2 | Analysis of lineages of transmissible neoplasia in cockles (*C. edule*). **a**, Representative *C. edule* (scale bar, 20 mm). **b**, **c**, Microsatellite loci were amplified from solid tissue (T) and haemocyte (H) DNA from normal (samples N1–N6) and diseased (M1–M3 and H1–H3) cockles. Products of a representative locus (CeATC1–5) are shown. Triangles mark samples corresponding to lineages 1 (red) and 2 (purple). **d**, Neighbour-joining phylogenetic tree based on nine microsatellite loci (see Source data). Bootstrap values over 50 are shown and scale bar is

of disease to *M. arenaria*²⁶, and injection of *M. trossulus* cancer cells into multiple bivalve species only resulted in engraftment in *M. trossulus*²⁷. Our finding that multiple cancer lineages are most often found to spread within the original host species is consistent with these previous experiments, and suggests that there may be species-specific restriction factors that prevent engraftment into divergent hosts.

In this context, our observation of cross-species transmission, a cancer from one species spreading through another, is particularly striking. They both belong to the Veneridae family (some studies suggest that *P. aureus* should belong to the *Venerupis* genus²⁸), and coexist in the same beds. This close relationship may have aided the transmission.

based on comparison across all 104 observed alleles. *mtCOI* SNPs unique to lineage 2 are marked. **e**, Maximum likelihood tree of *EF1α* (2725–4249 bp, spanning four introns), rooted at the midpoint, with bootstrap values above 50 and scale bar showing genetic distance. Letters 'a' and 'b' denote multiple alleles from heterozygous samples, and alleles from neoplastic lineages 1 and 2 are marked. **f, g**, Histology of cockles H3 and H2, representative of type A neoplastic cells. **h, i**, Histology of cockle H1, representative of type B neoplastic cells.

It is notable that despite ongoing surveillance of the bivalves of the Galician coast since the 1990s, only one *V. corrugata* has been identified as harbouring disseminated neoplasia (S. Darriba, personal communication), while the prevalence of disease in *P. aureus* is quite high¹² (12% of the individuals collected for the current study had high levels of neoplasia and 42% had some detectable level of disease). This would be explained if the cancer originated in *V. corrugata*, led to selective loss of susceptible animals, and left the current population of *V. corrugata* resistant to engraftment and disease. The spread of transmissible cancers and the evolutionary pressure to defend against them may be a strong and underappreciated selective force in the evolution of multicellular organisms.

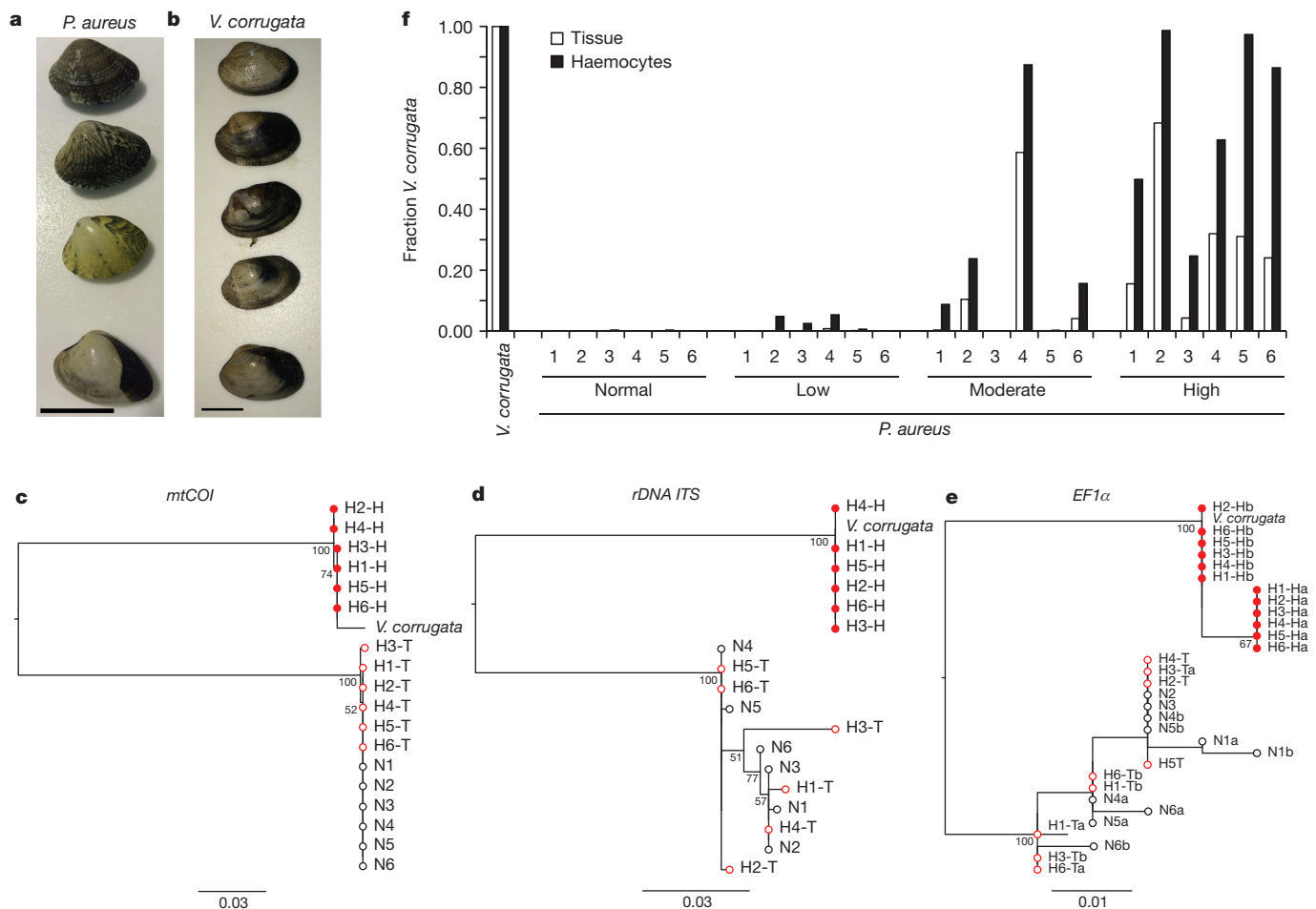


Figure 3 | Phylogenetic analysis of neoplastic cells in *P. aureus* and quantification of *V. corrugata* cell engraftment in normal and diseased animals. a, Representative *P. aureus* clams (scale bar, 20 mm). b, Representative *V. corrugata* clams (scale bar, 20 mm). c–e, Maximum likelihood trees of sequences from solid tissue (T) and haemocyte (H) DNA from normal (N1–N6) and highly diseased (H1–H6) *P. aureus* and normal *V. corrugata*, based on (c) *mtCOI* (658 bp), (d) *rDNA ITS* (454–373 bp), and (e) *EF1α* (148 bp, letters ‘a’ and ‘b’ denote multiple

alleles). Open black circles, tissue and haemocytes of normal *P. aureus*; open red circles, tissue of diseased *P. aureus*; filled red circles, haemocytes of diseased *P. aureus*; scale bars show genetic distance. f, Quantification of cancer cell engraftment using species-specific amplification of *EF1α* from haemocyte and tissue DNA of normal *V. corrugata*, normal *P. aureus*, and *P. aureus* diagnosed as having low, moderate, or high levels of disseminated neoplasia.

In summary, our findings show that transmission of cancers between individuals within a species may be more common than previously assumed, and that transmission can even occur between species. We now know of eight transmissible cancers in nature: one lineage in dogs, two lineages in Tasmanian devils, and five lineages circulating in four species of molluscs, including one example of cross-species transmission. Further studies may well reveal additional examples. The recent report of a cancer composed of transformed tapeworm cells in an immunocompromised AIDS patient²⁹ highlights the possibility that transmissible tumours could arise in humans. These transmissible cancers constitute a distinct class of infectious agent and show the remarkable ability of tumours to acquire new phenotypes that promote their own survival and propagation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 January; accepted 24 May 2016.

Published online 22 June 2016.

- Pearse, A. M. & Swift, K. Allograft theory: transmission of devil facial-tumour disease. *Nature* **439**, 549 (2006).
- Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A. & Weiss, R. A. Clonal origin and evolution of a transmissible cancer. *Cell* **126**, 477–487 (2006).

- Rebeck, C. A., Thomas, R., Breen, M., Leroi, A. M. & Burt, A. Origins and evolution of a transmissible cancer. *Evolution* **63**, 2340–2349 (2009).
- Metzger, M. J., Reinisch, C., Sherry, J. & Goff, S. P. Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* **161**, 255–263 (2015).
- Barber, B. J. Neoplastic diseases of commercially important marine bivalves. *Aquat. Living Resour.* **17**, 449–466 (2004).
- Carballal, M. J., Barber, B. J., Iglesias, D. & Villalba, A. Neoplastic diseases of marine bivalves. *J. Invertebr. Pathol.* **131**, 83–106 (2015).
- Moore, J. D., Elston, R. A., Drum, A. S. & Wilkinson, M. T. Alternate pathogenesis of systemic neoplasia in the bivalve mollusc *Mytilus*. *J. Invertebr. Pathol.* **58**, 231–243 (1991).
- Vassilenko, E. & Baldwin, S. A. Using flow cytometry to detect haemic neoplasia in mussels (*Mytilus trossulus*) from the Pacific Coast of Southern British Columbia, Canada. *J. Invertebr. Pathol.* **117**, 68–72 (2014).
- Vassilenko, E. I., Muttay, A. F., Schulte, P. M. & Baldwin, S. A. Variations in p53-like cDNA sequence are correlated with mussel haemic neoplasia: a potential molecular-level tool for biomonitoring. *Mutat. Res.* **701**, 145–152 (2010).
- Villalba, A., Carballal, M. J. & López, C. Disseminated neoplasia and large foci indicating heavy haemocytic infiltration in cockles *Cerastoderma edule* from Galicia (NW Spain). *Dis. Aquat. Organ.* **46**, 213–216 (2001).
- Díaz, S., Iglesias, D., Villalba, A. & Carballal, M. J. Long-term epidemiological study of disseminated neoplasia of cockles in Galicia (NW Spain): temporal patterns at individual and population levels, influence of environmental and cockle-based factors and lethality. *J. Fish Dis.* <http://dx.doi.org/10.1111/jfd.12436> (2016).
- Carballal, M. J., Iglesias, D., Díaz, S. & Villalba, A. Disseminated neoplasia in clams *Venerupis aurea* from Galicia (NW Spain): histopathology, ultrastructure and ploidy of the neoplastic cells, and comparison of diagnostic procedures. *J. Invertebr. Pathol.* **112**, 16–19 (2013).

13. Carballal, M. J., Iglesias, D., Santamarina, J., Ferro-Soto, B. & Villalba, A. Parasites and pathologic conditions of the cockle *Cerastoderma edule* populations of the coast of Galicia (NW Spain). *J. Invertebr. Pathol.* **78**, 87–97 (2001).
14. Iglesias, D. *Estudio Patológico de las Poblaciones de Berberecho Cerastoderma edule (L.) de Galicia*. Ph.D. thesis, Univ. Santiago do Compostela (2006).
15. Martínez, L., Arias, A., Méndez, J., Insua, A. & Freire, R. Development of twelve polymorphic microsatellite markers in the edible cockle *Cerastoderma edule* (Bivalvia: Cardiidae). *Conserv. Genet. Resour.* **1**, 107–109 (2009).
16. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
17. Rodzen, J. A. & May, B. Inheritance of microsatellite loci in the white sturgeon (*Acipenser transmontanus*). *Genome* **45**, 1064–1076 (2002).
18. Arriagada, G. *et al.* Activation of transcription and retrotransposition of a novel retroelement, Steamer, in neoplastic hemocytes of the mollusk *Mya arenaria*. *Proc. Natl Acad. Sci. USA* **111**, 14175–14180 (2014).
19. Pye, R. J. *et al.* A second transmissible cancer in Tasmanian devils. *Proc. Natl Acad. Sci. USA* **113**, 374–379 (2016).
20. Cockrill, J. M. & Beasley, J. N. Transmission of transmissible venereal tumor of the dog to the coyote. *Am. J. Vet. Res.* **40**, 409–410 (1979).
21. Samso, A. *Recherches Experimentales sur le Sarcome de Sticker*. Docteur ès Sciences Naturelles thesis, Univ. Paris (1965).
22. Sticker, A. Transplantables Rundzellensarkom des Hundes. *Z. Krebsforsch.* **4**, 227–314 (1906).
23. Wade, H. An experimental investigation of infective sarcoma of the dog, with a consideration of its relationship to cancer. *J. Pathol. Bacteriol.* **12**, 384–425 (1908).
24. Dungern, V. Zur Biologie des Rundzellensarkoms des Hundes. *Munch. med. Wochenschr.* No. 5, 238–239 (1912).
25. Strakova, A. & Murchison, E. P. The changing global distribution and prevalence of canine transmissible venereal tumour. *BMC Vet. Res.* **10**, 168 (2014).
26. Mateo, D. R., MacCallum, G. S. & Davidson, J. Field and laboratory transmission studies of haemic neoplasia in the soft-shell clam, *Mya arenaria*, from Atlantic Canada. *J. Fish Dis.* <http://dx.doi.org/10.1111/jfd.12426> (2015).
27. Kent, M. L., Wilkinson, M. T., Drum, A. S. & Elston, R. A. Failure of transmission of hemic neoplasia of bay mussels, *Mytilus trossulus*, to other bivalve species. *J. Invertebr. Pathol.* **57**, 435–436 (1991).
28. Canapa, A., Schiaparelli, S., Marota, I. & Barucca, M. Molecular data from the 16S rRNA gene for the phylogeny of Veneridae (Mollusca: Bivalvia). *Mar. Biol.* **142**, 1125–1130 (2003).
29. Muehlenbachs, A. *et al.* Malignant transformation of *Hymenolepis nana* in a human host. *N. Engl. J. Med.* **373**, 1845–1852 (2015).

Acknowledgements M.J.M. and S.P.G. were supported by the Howard Hughes Medical Institute and Training Grant T32 CA009503. D.I., M.J.C., and A.V. were supported by the Consellería do Mar da Xunta de Galicia, through the project PGIDIT-CIMA 13/03. We thank J. Ausió for help in collection of *M. trossulus* from Vancouver Island.

Author Contributions M.J.M. and S.P.G. wrote the manuscript. M.J.M. conducted molecular analyses. A.F.M. and S.A.B. collected and diagnosed *M. trossulus* from West Vancouver. J.S. and C.R. collected and diagnosed *M. trossulus* from Vancouver Island. D.I., M.J.C., and A.V. collected and diagnosed *C. edule* and *P. aureus*. M.J.C. produced micrographs of *C. edule* neoplastic haemocytes, and D.I. conducted morphometric analysis.

Author Information Sequences generated in this work have been deposited in GenBank under accession numbers KX018521–KX018605. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P.G. (spg1@cumc.columbia.edu).

Reviewer Information Nature thanks E. Murchison, S. O'Brien, G. De Vico, R. A. Weiss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to diagnosis during sequencing and morphology assessment.

Collection and diagnosis of *M. trossulus*. Mussel (*M. trossulus*) specimens 5–6 cm long were collected from the intertidal zone at low tide (noon) on 18 April 2015 at Copper Beach (49° 22' 41" N, 123° 16' 44" W, West Vancouver, British Columbia, Canada). They were transported to the laboratory in aerated seawater from Copper Beach. In the laboratory, 0.5–1.0 ml haemolymph was removed from the posterior adductor muscle. For each individual, one drop of haemolymph was placed on a poly-L-lysine-coated slide and let sit for 10–15 min to allow the cells to spread and attach to the slide. Thereafter, the slide was viewed under a Zeiss Axiostar light microscope at $\times 40$ magnification. Normal (non-neoplastic) specimens were those with greater than 90% cells with normal appearance: that is, agranular or granular haemocytes with spread pseudopodia³⁰. Fully leukaemic (diseased) specimens were those with prolific amounts (>90%) of round, non-adherent cells. Haemolymph was added to 1.5 ml Eppendorf tubes and spun at 900g for 3 min. After the supernatant was withdrawn, the cells were re-suspended in absolute ethanol before shipping. Excised tissues (mantle, foot, and gills) were preserved in absolute ethanol before shipping. Of 28 individuals collected from West Vancouver, two had high levels of neoplastic disease.

A second set of *M. trossulus* samples were collected from Esquimalt, Vancouver Island, British Columbia, Canada. Haemolymph was extracted and cell morphology was used for diagnosis as above. Haemocyte and solid tissue samples were fixed in ethanol before DNA extraction. Of 250 individuals collected from Esquimalt, 9 were scored as potentially moderately or highly diseased, with 7 samples available for analysis.

Collection and diagnosis of *C. edule* and *P. aureus*. Cockles (*C. edule*) were collected from an intertidal bed named O Sarrido (42° 30' N, 8° 49' W) and golden carpet shell clams (*P. aureus*) and pullet shell clams (*V. corrugata*) were collected from a subtidal bed named O Bohido (42° 32' N, 8° 51' W); both shellfish beds are located in the ria of Arousa (Galicia, northwest Spain). Once in the laboratory of Centro de Investigaciones Mariñas, each cockle and clam was notched through the shell margin close to the posterior adductor muscle and haemolymph (as much as possible) was collected from the posterior adductor muscle using a 2-ml syringe with a 21-gauge needle. A small quantity (about 100 µl) of haemolymph was used to produce a cell monolayer onto a slide by cytocentrifugation (92g, 5 min, 4°C), which was fixed, stained with a Hemacolor (Merck) kit, and examined with light microscopy for diagnosis of disseminated neoplasia; the remaining haemolymph was preserved in absolute ethanol for molecular analysis. After haemolymph collection, molluscs were shucked and a small piece of mantle (about 5 mm \times 5 mm) was removed and preserved in absolute ethanol for molecular analysis; additionally, an approximately 5-mm-thick section of meat, containing gills, visceral mass, foot, and mantle lobes, was fixed in Davidson's solution and embedded in paraffin. Sections of 5 µm thickness were stained with Harris's haematoxylin and eosin³¹. Histological sections were examined under light microscopy for histopathological analysis.

Morphometric analysis was conducted on histological sections of cockle (*C. edule*) samples. Neoplastic cells (both A and B types) had a unique morphology, and were clearly distinguished from normal cells, with much larger overall size, and much larger nucleus. The longest diameters of the cell and the nucleus of at least ten neoplastic cells for each individual were measured by direct examination of histological sections with light microscopy using a reticle. Two-tailed *t*-tests were used for comparisons of morphometric data from different individuals and types.

Cockles (*C. edule*) and golden carpet shell clams (*P. aureus*) were ranked according to a scale of disease severity based on haemolymph diagnosis: non-affected (N0, or N); low severity (N1, or L), when individuals showed proportions of neoplastic cells lower than 15% in the haemolymph cell monolayers; moderate severity (N2, or M), when the proportion ranged from 15% to 75%; and high severity (N3, or H), when the proportion was higher than 75% (ref. 32). Seventy-four golden carpet shell clams (*P. aureus*) were collected with 12 diagnosed with light, 10 with moderate, and 9 with heavy neoplasia. About 150 cockles (*C. edule*) were collected and a subset was analysed both for disease and for morphological type of neoplastic cells. Of the 30 in this subset, two had type A neoplastic cells (one light and the other moderate) and one had low levels of type B neoplastic cells. We have also collected and analysed hundreds of pullet clams (*V. corrugata*) from multiple beds throughout Galicia from 1988 to the present, including about 100 from the same bed in which the samples of *P. aureus* were collected for this study (O Bohido). So far, we have not observed any *V. corrugata* samples with neoplastic disease. One *V. corrugata* individual from a different location has been identified by a different researcher as harbouring disseminated neoplasia, with a different morphology than that observed in *P. aureus* neoplasia (S. Darriba, personal communication).

DNA extraction. DNA was extracted from ethanol-fixed haemocytes using DNeasy Blood and Tissue Kit (Qiagen). DNA extraction of tissues used the same kit, but included an additional step to reduce the amount of PCR-inhibiting polysaccharides. After tissue lysis, 63 µl of buffer P3 was added to lysate and allowed to precipitate for 5 min. Lysate was spun 10 min at full speed at 4°C, and the resulting supernatant was mixed with buffer AL for 10 min at 37°C, then mixed with ethanol and added to the column, continuing with the standard protocol.

mtCOI, EF1 α , ITS, Steamer-like PCR. Primers and annealing temperatures used are listed in Extended Data Table 2. PfuUltra II Fusion HS DNA Polymerase (Agilent) was used to amplify 10 ng of genomic DNA for 35 cycles. Extension was at 72°C for 15 s for *mtCOI* for all bivalves³³. For *P. aureus*, *ITS* was amplified (using primers modified from refs 34, 35) with extension for 30 s. Despite multiple copies of *ITS* in genomic DNA, a single sequence was observed for each normal *P. aureus* and *V. corrugata*, and a single pair of host/neoplasm sequences was observed in each diseased individual. PCR for *EF1 α* followed the same program, with extension for 20 s in mussels (*M. trossulus*) and clams (*P. aureus* and *V. corrugata*) and 1 min 30 s in cockles (*C. edule*). For amplification of *Steamer*-like elements, a degenerate primer pair was designed to match conserved regions in reverse transcriptase and integrase of the LTR-retrotransposon, *Steamer*, and 10 ng of DNA was amplified for 35 cycles, annealing at 55°C for 20 s and extending for 1 min at 72°C. PCR products were directly sequenced. When multiple alleles could not be resolved by direct sequencing, PCR products were cloned using the Zero Blunt TOPO Kit (Invitrogen), and at least six colonies were sequenced. Sequences were aligned with Clustal W, with some manual adjustment. Primer-binding regions were excluded from analysis and are not included in the sizes listed as sequenced. Maximum likelihood phylogenetic trees were generated using PhyML 3.0 (ref. 36), using the HKY85 substitution model, with 100 bootstrap replicates, treating gaps in the alignment as missing data. Trees were visualized using FigTree version 1.4, with addition of markers at the branch termini. Each phylogenetic tree based on alignment of sequence at a single locus (Figs 1c, 2e and 3b–d) includes a scale bar that shows genetic distance (0–1) based on the frequency of nucleotide divergence. All sequences are available in GenBank (accession numbers KX018521–KX018605).

Analysis of microsatellite loci from *C. edule*. Microsatellites were amplified using primers for 12 loci, reported previously to be polymorphic in cockle (*C. edule*) populations¹⁵. Of the 12 primer pairs, 9 pairs amplified products in all cockle DNA samples, with 1–2 alleles observed in normal samples and 1–4 neoplastic alleles observed. These were used in all further analyses, with fluorescent modifications on the 5' end of the forward primers as listed (Extended Data Table 2). KOD polymerase (Millipore) was used to prevent ambiguity of untemplated addition of A residues by *Taq* polymerase. Products were run on an agarose gel for visualization. Allele sizes were identified to single-base precision by fragment analysis using fluorescent primers (6-FAM, PET, NED, and VIC) using a 3730xl Genetic Analyzer with the LIZ-500 size standard (Applied Biosystems, operated by Genewiz). Peaks were called using Peak Scanner 2.0 (Applied Biosystems), rounding sizes to whole bases, with some manual adjustment to keep alleles of the same size together.

We used the R package *Poppr*¹⁶ to generate distance matrices and phylogenetic trees using Bruvo's method with the infinite alleles model. The value of *c* (repeat size) was set to 0.00001 to calculate distances using the band sharing model, in which alleles are either identical sizes (distance = 0) or non-identical (distance = 1). Each allele was coded as a dominant marker¹⁷, with a single variable for each allele observed at each position (104 total markers from 9 loci). For each allele (a single size at a single locus), a sample is observed to be present, absent, or the information missing. For example, a sample with sizes 195 and 199 at locus A would be marked as 'present' for marker 'A-195' and 'A-199', but 'absent' for all other sizes at that locus (like 'A-203' and 'A-207'). This analysis method ignores the uncertainty in the copy number of each locus in aneuploid cancer cells and allows for ambiguity at particular alleles. In cases where the haemocyte and tissue genotypes could be clearly differentiated, all alleles were analysed (H1 and H3). In some cases (M1, M2, M3, and H2), cancer-specific alleles could be detected in the haemocyte samples, but represented less than 50% of the total haemocyte DNA, and the host tissue alleles obscured one or two positions (depending on whether the normal genotype was homozygous or heterozygous at that locus). In these cases, since the tissue allele obscured a potential neoplastic allele, that size was coded as 'missing' for the haemocyte genotype. The total distance between any two samples is calculated as the average of the pairwise differences at all allele sizes observed (104 pairwise comparisons if no data are missing). In pairwise comparisons for each allele, two samples either have a distance of 0 (meaning that the particular allele size is either present in both samples or absent in both samples) or a distance of 1 (meaning that the particular allele size is present in one sample but absent in the other). If data are missing for one or both, then that specific comparison is dropped with no contribution to the total distance between the two samples. The genetic distances calculated are based on the alleles observed at the nine loci, so the absolute value

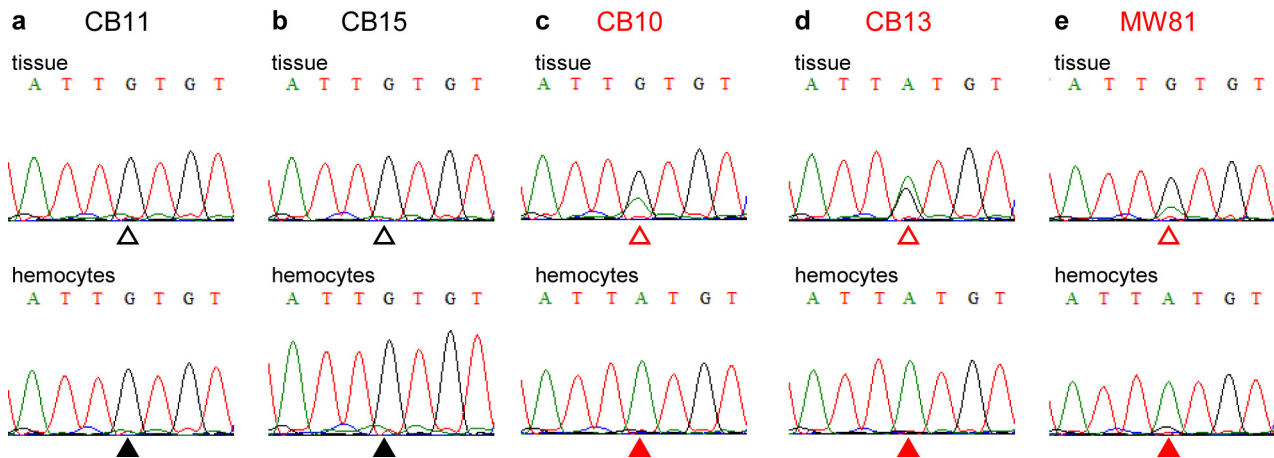
of the scale itself is therefore dependent on the number of observed alleles that are included in the analysis. Source data for the generation of the phylogenetic tree lists all observed alleles coded as present (1), absent (2), or missing (0) for each sample.

As an alternative analysis method, the data were analysed using Bruvo's method as polyploid data with nine loci, using only the individuals that could be confidently identified. In this alternative analysis, the topology of the resulting tree was not different—all nodes with bootstrap values above 50 were maintained. In this analysis, the genetic distance between the representatives of lineage 1 (H3) and lineage 2 (H1) was 0.676, and the distances to the closest normal sample were 0.595 (H3 to N5) and 0.648 (H1 to N1 and N2).

qPCR. Species-specific qPCR of the *EF1 α* locus in *P. aureus* and *V. corrugata* was done using FastStart Universal SYBR Green Master Mix (Roche). Primers were designed to amplify the same region, with the 3' end of both forward and reverse primers on sites that differ between the two species (Extended Data Table 2). Standard curves for each primer set were generated using two control plasmids (one containing the *P. aureus* *EF1 α* fragment and *P. aureus* *SLE* fragment, the other containing the fragments from *V. corrugata*). *EF1 α* fragments were amplified using conserved primers (Extended Data Table 2). Fragments were cloned with a Zero Blunt TOPO Kit (Invitrogen), and plasmids were linearized with NotI before qPCR. Standard curve samples (10^4 – 10^2 copies per reaction) and experimental samples (2.5 ng per reaction) were done in triplicate. No amplification was detected with the *P. aureus* primers on the *V. corrugata* plasmid or with the *V. corrugata* primers on *P. aureus* plasmid (using up to 10^7 copies per reaction). The fraction of *V. corrugata* in each sample was calculated as the copy number using *V. corrugata*-specific primers divided by copy number of *P. aureus*- plus *V. corrugata*-specific amplification.

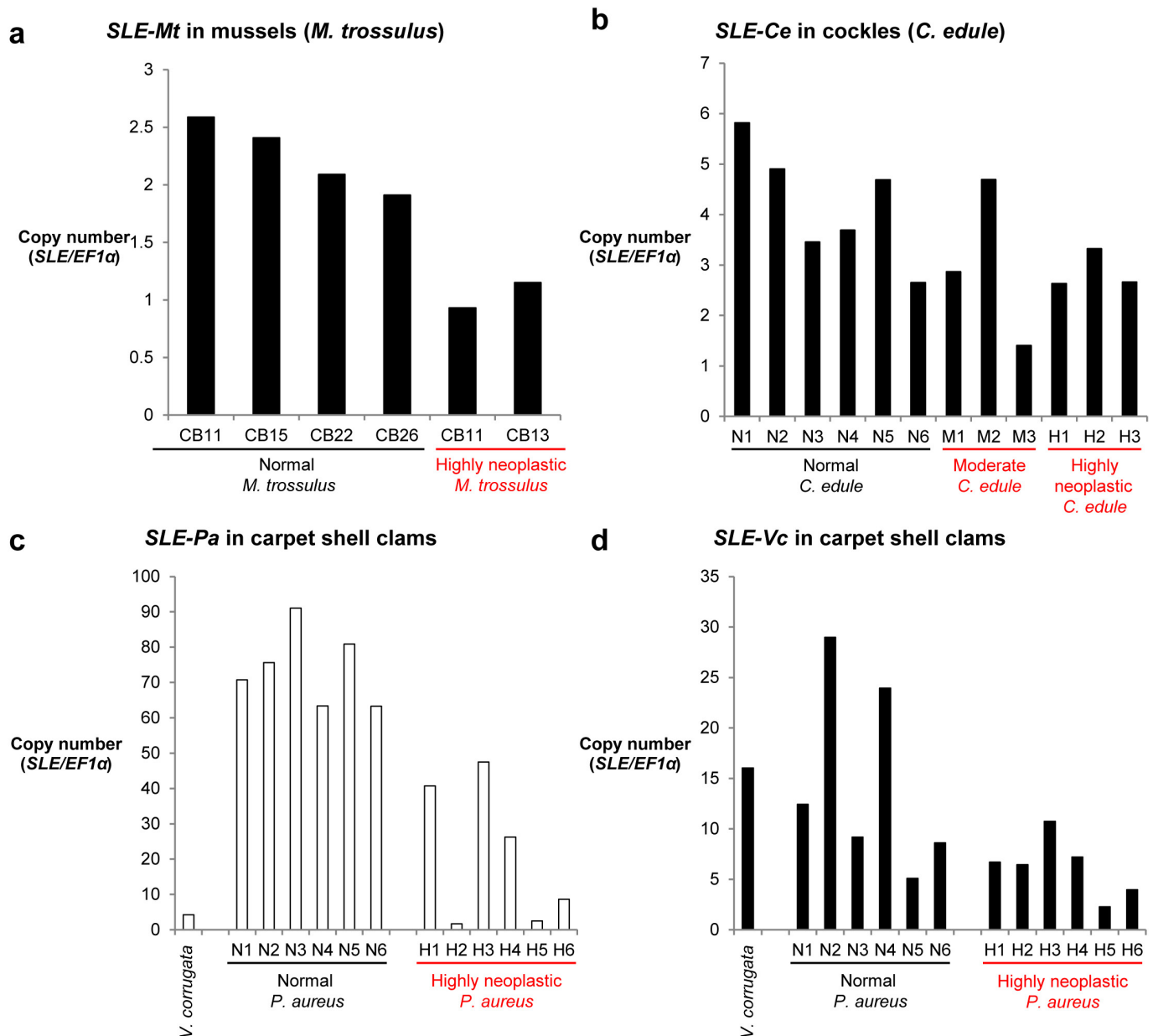
Quantification of *Steamer*-like elements in the three species was performed by the same method, using primers in the RT-IN region of the retrotransposon (on the basis of sequence obtained through amplification using degenerate primers) and control primers amplifying a region in *EF1 α* . All primers used to amplify *SLEs* and *EF1 α* from genomic DNA for cloning into standards and primers for qPCR are listed in Extended Data Table 2. For each species, a single control plasmid containing both the *EF1 α* and *SLE* fragments was used.

30. Muttray, A. F., Schulte, P. M. & Baldwin, S. A. Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia. *Mar. Environ. Res.* **66**, 412–421 (2008).
31. Howard, D., Lewis, E., Keller, B. & Smith, C. *NOAA Technical Memorandum NOS NCCOS Vol. 5* (NOAA/National Centers for Coastal Ocean Science, 2004).
32. Díaz, S., Cao, A., Villalba, A. & Carballal, M. J. Expression of mutant protein p53 and Hsp70 and Hsp90 chaperones in cockles *Cerastoderma edule* affected by neoplasia. *Dis. Aquat. Organ.* **90**, 215–222 (2010).
33. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
34. Oliverio, M. & Mariottini, P. Contrasting morphological and molecular variation in *Coralliophila meyendorffii* (Muricidae, Coralliophilinae). *J. Molluscan Stud.* **67**, 243–245 (2001).
35. Salvi, D. & Mariottini, P. Molecular phylogenetics in 2D: ITS2 rRNA evolution and sequence-structure barcode from Veneridae to Bivalvia. *Mol. Phylogenet. Evol.* **65**, 792–798 (2012).
36. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).



Extended Data Figure 1 | Analysis of *mtCOI* amplified from tissue and haemocyte DNA of normal and diseased mussels (*M. trossulus*). A partial region of the *mtCOI* gene was amplified from genomic DNA of solid tissue and haemocytes from mussels (*M. trossulus*) and directly sequenced (Fig. 1b). Trace images show a region flanking a representative SNP marked with an open triangle (tissue) or closed triangle

(haemocytes). **a, b**, In normal mussels, tissue and haemocyte alleles match (with G496 at all positions). **c–e**, In mussels with disseminated neoplasia, the tissue and haemocyte alleles are different. Neoplastic haemocytes have A at position 496 and G in tissue, with some A observable in tissue, probably because of infiltration of neoplastic haemocytes.



Extended Data Figure 2 | Quantification of Steamer-like element genomic copy number in mussels (*M. trossulus*), cockles (*C. edule*), and golden carpet shell clams (*P. aureus*). a–d, Fragments from the SLE reverse transcriptase region and *EF1α* genes were cloned from each species. Haploid copy numbers of Steamer-like elements (SLE) were quantified by determining the ratio of SLE/*EF1α* in genomic DNA from haemocytes. Single species-specific SLEs were analysed in (a) mussels

(*M. trossulus*) and (b) cockles (*C. edule*). c, d, In golden carpet shell clams, one SLE (SLE-Pa) was cloned from a normal *P. aureus* (clam N2) and a different one (SLE-Vc) was cloned from neoplastic cells (clam H2). Both SLEs could be found in both species, and qPCR analysis confirmed that SLE-Pa is more highly amplified in *P. aureus* and has fewer copies in *V. corrugata* and in the neoplastic cells derived from *V. corrugata*.

Extended Data Table 1 | Morphometric analysis of type A and type B cockle (*C. edule*) neoplasia

Cockle ID	Type	Cells counted (N)	Cell diameter ± SD (μm) [*]	Nuclear diameter ± SD (μm) [†]	Nuclear/cell ratio ± SD [‡]
M1	A	15	9.0 ± 1.2	6.7 ± 1.0	0.75 ± 0.09
H2	A	15	8.7 ± 0.9	6.6 ± 0.8	0.76 ± 0.06
H3	A	10	8.5 ± 1.1	6.5 ± 1.0	0.77 ± 0.07
total	A	40	8.8 ± 1.0	6.6 ± 0.9	0.76 ± 0.07
M2	B	10	7.7 ± 0.7	5.5 ± 0.5	0.72 ± 0.07
M3	B	15	6.7 ± 1.2	4.5 ± 0.6	0.70 ± 0.15
H1	B	15	7.3 ± 1.0	5.4 ± 0.8	0.75 ± 0.09
total	B	40	7.2 ± 1.1	5.1 ± 0.8	0.72 ± 0.11

^{*}Cell diameter is statistically different between types A and B ($P < 0.0001$). All tests are two-tailed *t*-tests.

[†]Nuclear diameter is statistically different between types A and B ($P < 0.0001$); for all pairwise comparisons between type A and type B individuals, $P < 0.05$.

[‡]Ratio of nuclear diameter to cell diameter is not significantly different between types.

Extended Data Table 2 | Primers used in PCR and qPCR

Target	Forward Primer		Reverse Primer		Temp
mtCOI	LCO1490	GGTCAACAAATCATAAAGATATTGG	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	50°C
ITS	its-3d	gcgtcgatgaagagcgca	its-4r	agtttttttctcccgctta	55°C
EF1α					
Mussel (<i>M. trossulus</i>)	consEF1-F1	ACCATTGATATTGCTYTNITGGAA	MtEF1-R2	ACAATCAAAATGGCACAATC	50°C
Cockle (<i>C. edule</i>)	CeEF1-F2	TGGTATCACCATCGATATTGC	CeEF1-R3	CCGTTTCGGATCTCTACAGG	55°C
<i>P. aureus</i> & <i>V. corrugata</i>	consVEF1F3b	AGGAACCTCTCAAGCTGAYTG	consVEF1R3b	AGCAATTACCTCGCTGTANGG	55°C
Cockle (<i>C. edule</i>) Microsatellite Loci					
	CeATC1-5F-6FAM	cgttctaccggcatatgtcac	CeATC1-5R	caccttcaccactagaagaaaa	60°C
	CeATC1-22F-VIC	caaacctgaccgggtttatt	CeATC1-22R	tgactccactttttcagttcca	60°C
	CeATC1-36F-NED	gacatgacaaacaggcctca	CeATC1-36R	tggcctgggtcttatttccac	60°C
	CeATC1-52F-PET	aatctgattttgccacctct	CeATC1-52R	agctcataggagttgtatacgtaag	55°C
	CeATC1-54F-6FAM	tacaaggccgagaaactgct	CeATC1-54R	caatgactgccaaatgagga	55°C
	CeATC2-4F-VIC	tggaaatgcattcattgagc	CeATC2-4R	cggattgcgttctttgatct	55°C
	CeATC2-11F-NED	tgggtgtgcaattagatgcttg	CeATC2-11R	taggctcgagaaagatgggt	60°C
	CeATC2-34F-6FAM	gccatagaggccaccctatt	CeATC2-34R	gggctgacaagatttgacatt	60°C
	CeATC2-46F-NED	accaaggcagatatcgatcc	CeATC2-46R	tccagttttaacgcactctga	55°C
Cloning qPCR Standards					
EF1α					
Mussel (<i>M. trossulus</i>)	consEF1F1	ACCATTGATATTGCTYTNITGGAA	consEF1R2	ACGTTGAAACCAACRTTRTC	50°C
Cockle (<i>C. edule</i>)	CeEF1-F1	TGGAACAACTGAAGCCGGA	CeEF1-R3	CCGTTTCGGATCTCTACAGG	50°C
<i>P. aureus</i> & <i>V. corrugata</i>	consVEF1F3b	AGGAACCTCTCAAGCTGAYTG	consVEF1R3b	AGCAATTACCTCGCTGTANGG	55°C
SLE					
Mussel (<i>M. trossulus</i>)	DHKPL-F1	TTGAAAGCGACCACAARCCNYT	PXRPW-R1	TTCCGACTTTGGCCCANGGNC	55°C
Cockle (<i>C. edule</i>)	LWW-F1	GGTATCCAGCCTAGTNGTNGT	PXRPW-R1	TTCCGACTTTGGCCCANGGNC	55°C
<i>P. aureus</i> & <i>V. corrugata</i>	DHKPL-F1	TTGAAAGCGACCACAARCCNYT	PXRPW-VaR1	CTGCAATTTCTGCCANGGNC	55°C
qPCR	Forward Primer		Reverse Primer		Control Plasmids
EF1α					
Mussel (<i>M. trossulus</i>)	MtEF1qF2	TAGGTATTGGAACAGTGCCAGT	MtEF1qR1	AGACTCGTGGTGCAATTTCTAC	pMtEF1-SLE4
Cockle (<i>C. edule</i>)	CeEF1qF2	CCAGTGGGCCGAGTTGAGAC	CeEF1qR2	CTTCAGTGGTCACGTGGCAG	pCockleSLE-EF1
<i>P. aureus</i>	VaEF1qF1	AAAGAATGGACAGACCAGAGAG	VaEF1qR1	TGCTTCACACCAATGTGAAA	pVaN2-EF1SLE
<i>V. corrugata</i>	VcEF1qF1	CAAGAATGGACAGACAAGAGAA	VcEF1qR1	TGTTTCACACCAAGGTGAAG	pVaH2-EF1SLE
SLE					
<i>SLE-Mt</i>	MtSLEqF2	TGGACTACTTTACAAAAGCCAACG	MtSLEqR1	GCTTCGTGCAATTTCACTAACAT	pMtEF1-SLE4
<i>SLE-Ce</i>	cockleRT-F2	GCAGTGTGCAATGTCCCAGT	cockleRT-R2	GACCTGTGTCCGAGGCATCA	pCockleSLE-EF1
<i>SLE-Pa</i>	VaNSLEqF1	TAGAGACGAATTTGCCGTGGG	VaNSLEqR4	CGTGCTCGCTGTAAGCATTT	pVaN2-EF1SLE
<i>SLE-Vc</i>	VaHSLEqF1	TAGAGATGAACCTCTCTTTGA	VaHSLEqR1	TGCACGGCTTAATGTTTGGGA	pVaH2-EF1SLE

Mitochondrial unfolded protein response controls matrix pre-RNA processing and translation

Christian Münch¹ & J. Wade Harper¹

The mitochondrial matrix is unique in that it must integrate the folding and assembly of proteins derived from the nuclear and mitochondrial genomes. In *Caenorhabditis elegans*, the mitochondrial unfolded protein response (UPR^{mt}) senses matrix protein misfolding and induces a program of nuclear gene expression, including mitochondrial chaperonins, to promote mitochondrial proteostasis^{1–3}. While misfolded mitochondrial-matrix-localized ornithine transcarbamylase induces chaperonin expression^{4–6}, our understanding of mammalian UPR^{mt} is rudimentary⁷, reflecting a lack of acute triggers for UPR^{mt} activation. This limitation has prevented analysis of the cellular responses to matrix protein misfolding and the effects of UPR^{mt} on mitochondrial translation to control protein folding loads. Here we combine pharmacological inhibitors of matrix-localized HSP90/TRAP1 (ref. 8) or LON protease⁹, which promote chaperonin expression, with global transcriptional and proteomic analysis to reveal an extensive and acute response of human cells to UPR^{mt}. This response encompasses widespread induction of nuclear genes, including matrix-localized proteins involved in folding, pre-RNA processing and translation. Functional studies revealed rapid but reversible translation inhibition in mitochondria occurring concurrently with defects in pre-RNA processing caused by transcriptional repression and LON-dependent turnover of the mitochondrial pre-RNA processing nuclease MRPP3 (ref. 10). This study reveals that acute mitochondrial protein folding stress activates both increased chaperone availability within the matrix and reduced matrix-localized protein synthesis through translational inhibition, and provides a framework for further dissection of mammalian UPR^{mt}.

Protein folding homeostasis is central to cell fitness. Protein unfolding in the endoplasmic reticulum (ER) promotes transcriptional induction of ER-associated chaperones to facilitate folding and inhibits translation to further reduce the folding load¹¹. In contrast, mechanisms underlying the response to protein misfolding in other organelles, including mitochondria, are poorly understood. The mitochondrial matrix folding machinery consists of chaperonins HSPD1/HSP60 and HSPE1/HSP10, and chaperones including the HSP90 paralogue TRAP1 and mtHSP70. This machinery assists in the folding of matrix-localized nuclear-encoded proteins, and their assembly with 13 respiratory chain proteins encoded by the mitochondrial genome (mtDNA)¹². The balance between folding load and chaperone abundance is controlled, in part, by the UPR^{mt}. In *C. elegans*, genetic UPR^{mt} activation promotes nuclear localization of the ATF5-1 transcription factor to induce expression of mitochondrial chaperonins, thereby enhancing matrix folding capacity^{1–3}. While earlier work revealed that enforced expression of misfolded ornithine transcarbamylase in HeLa cells induced *HSPD1* and *HSPE1* expression^{4–6}, our understanding of UPR^{mt} in human cells is limited.

Cellular stress responses such as UPR^{er} are typically fast acting as a result of rapid sensing of protein folding stress, but prolonged activation can produce confounding effects such as cell death¹³. We therefore examined whether gamitrinib-triphenylphosphonium (GTPP)—a

specific inhibitor of the matrix HSP90 chaperone TRAP1 known to cause protein misfolding in this compartment^{8,14}—would promote acute transcription of *HSPD1* and *HSPE1* as readout of UPR^{mt} induction in HeLa cells. Acute GTPP treatment (6 h) induced UPR^{mt} as assessed by quantitative PCR (qPCR) for *HSPD1* and *HSPE1* (Extended Data Fig. 1a) with a dynamic range (approximately twofold) similar to that seen with genetic UPR^{mt} induction in *C. elegans*¹. *HSPD1* and *HSPE1* are among the most abundant messenger RNAs (mRNAs) in untreated cells (top 2nd percentile), explaining their limited dynamic range upon UPR^{mt} (Supplementary Table 1). GTPP treatment did not affect cell viability, mitochondrial membrane potential, ATP levels or respiratory chain architecture (Extended Data Fig. 1b–e). Longer (24 h) incubations with GTPP result in cell death⁸. Consistent with TRAP1 being the causal target for GTPP-dependent chaperonin induction, TRAP1 RNA interference (RNAi) also induced *HSPD1* by qPCR (Extended Data Fig. 1f).

C/EBP homologous protein (CHOP), a broadly acting transcription factor, is induced via UPR^{er} and the integrated stress response via the ATF4 transcription factor¹¹. CHOP is also induced during UPR^{mt} (refs 4, 5) and oxidative stress¹⁵, but the mechanisms underlying CHOP activation in UPR^{mt} and its relationship between UPR^{er} and integrated stress response upstream signalling remained unclear. Strikingly, we found that GTPP, but not the UPR^{er} activator tunicamycin, respiratory chain inhibitors or mitochondrial membrane decouplers, activated *HSPD1* expression (Fig. 1a and Extended Data Fig. 2a). GTPP also activated *ATF4* and *CHOP*, but unlike tunicamycin, did not induce *BIP*, indicating that GTPP does not activate canonical UPR^{er} (Fig. 1b, c and Extended Data Fig. 2a, b). We also found that individual depletion of the four known EIF2A kinases involved in integrated stress response signalling (GCN2, HRI, PERK and PKR)¹¹ had no effect on *CHOP* induction by GTPP (Extended Data Fig. 2c–f), suggesting that induction of *ATF4* and *CHOP* by UPR^{mt} occurs through a pathway independent of individual integrated stress response kinases⁵ (Extended Data Fig. 2b). Taken together, these data indicate that GTPP induces UPR^{mt} through a pathway distinct from known ER and mitochondrial stress pathways (Extended Data Fig. 2b).

To globally examine the mammalian UPR^{mt} transcriptional response, we treated HeLa cells with GTPP for 6 h and performed RNA sequencing (RNA-seq) (Fig. 1d, e, Extended Data Fig. 3a–b and Supplementary Table 1). In parallel, we determined RNA-seq profiles upon treatment of cells with CDDO, an inhibitor of matrix protease LON (Fig. 1d). CDDO rapidly induces mitochondrial protein misfolding⁹ and induced *HSPD1* expression, consistent with UPR^{mt} induction (Extended Data Fig. 3c). From 968 (GTPP) and 1,029 (CDDO) transcripts whose abundance changed significantly ($P \leq 0.05$, $\log_2 \geq \pm 0.6$), 627 were shared between the two different treatments with 337 and 290 downregulated and upregulated transcripts, respectively, including *HSPD1* and *HSPE1*, and *CHOP* (Fig. 1d–f and Extended Data Fig. 3d, e). Importantly, changes in transcription with GTPP treatment were distinct from changes previously reported with 17-AAG¹⁶, a derivative of GTPP that inhibits cytoplasmic and nuclear HSP90 (Extended Data Fig. 3e), indicating that inhibition of non-mitochondrial HSP90 is unlikely to account for

¹Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

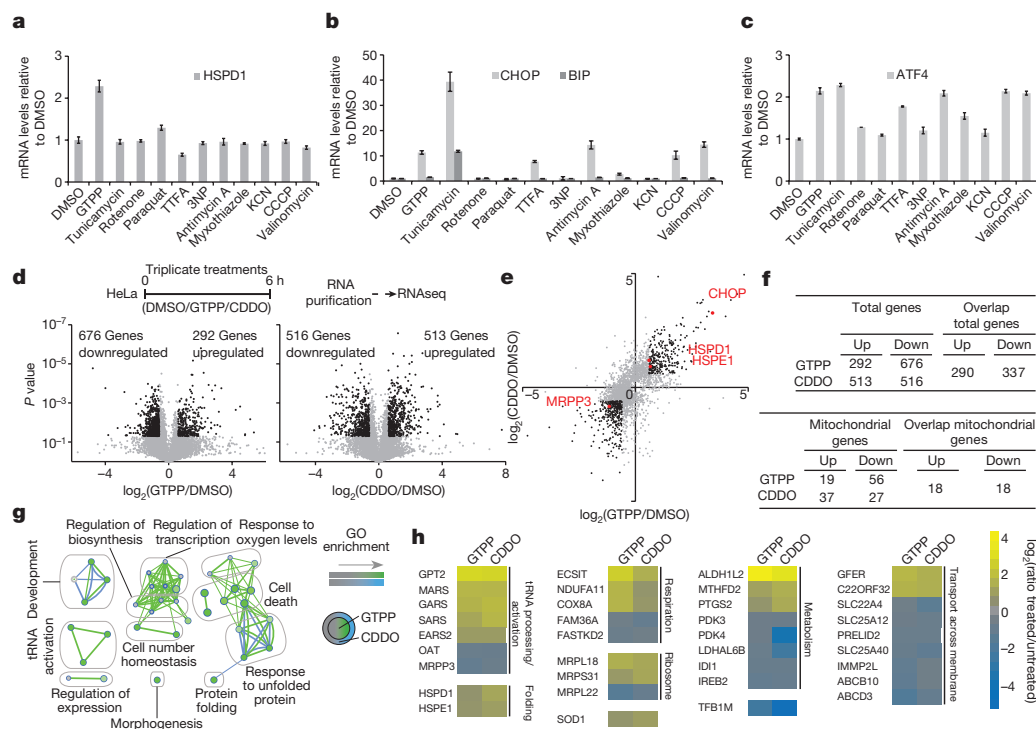


Figure 1 | Global analysis of transcriptional responses to UPR^{mt} induction. **a–c**, qPCR of *HSPD1* (**a**), *CHOP* and *BIP* (**b**) or *ATF4* (**c**) mRNA in HeLa cells with or without the indicated treatments (mean of levels relative to untreated \pm s.d.; $n = 3$ biological replicates). **d**, Experimental design (top). Volcano plot showing fold changes versus P values for the analysed transcriptome of cells treated with GTPP (bottom left) or CDDO (bottom right). Transcripts significantly changing upon UPR^{mt} induction ($P \leq 0.05$, changes $\geq \log_2 \pm 0.6$) are represented by black dots. **e**, Correlation of ratios of transcripts changing upon GTPP or CDDO treatment. Black dots, $P \leq 0.05$, changes $\geq \log_2 \pm 0.6$; red dots, genes of interest. **f**, Summary of altered transcripts. **g, h**, Gene ontology (GO) enrichment map (**g**) and heat map (**h**) of overlapping mitochondrial transcripts altered by both GTPP and/or CDDO.

the transcriptional response with GTPP. Gene ontology enrichment analysis confirmed extensive overlap in the transcriptional responses, with all gene ontology clusters representing transcripts altered with both treatments (Fig. 1g and Supplementary Table 2). As expected, gene ontology terms showed enrichment for protein folding genes, consistent with UPR^{mt} induction, but also included transfer RNA (tRNA) processing and activation. Among the nuclear genes with correlated changes in transcription, 36 encode proteins known to localize in mitochondria (Fig. 1h and Supplementary Table 1). Promoter analysis of genes regulated by UPR^{mt} induction showed enrichment of CHOP and ATF4 promoter recognition sequences, as well as two ‘mitochondrial UPR Response Element’ (MURE1 and MURE2) promoter elements⁶ ($P < 0.0001$; Extended Data Fig. 4 and Supplementary Table 3). This analysis therefore revealed a specific nuclear response to UPR^{mt} that is anticipated to promote homeostasis of protein folding within mitochondria.

We then applied MultiNotch proteomics¹⁷ (Extended Data Fig. 5a) to purified mitochondria to quantify acute changes in the mitochondrial proteome upon GTPP treatment using untreated cells or cells treated with the mitochondrial uncoupler CCCP (carbonyl cyanide-*m*-chlorophenyl hydrazone) as controls (Fig. 2a and Supplementary Table 4)¹⁷. From

606 mitochondrial proteins quantified (442 with 2 or more peptides), 61 displayed significant changes in abundance 6 h after GTPP treatment compared with control or CCCP-treated cells, including HSPD1 and HSPE1, which increased as expected (Fig. 2a, b and Extended Data Fig. 5b). Furthermore, proteins involved in respiration, transcription, tRNA processing and protein quality control, among others, were found to be regulated (Fig. 2c). In contrast, levels of the mitochondrial ribosome and respiratory chain complexes were not significantly altered (Extended Data Fig. 5c, d), consistent with their long half-lives¹⁸. Strikingly, the abundance of the mitochondrial matrix protein MRPP3 was reduced at both the transcriptional and protein level (Fig. 1e, h and Fig. 2b–d). MRPP3 is the catalytic subunit of the RNA-free mitochondrial RNase P complex, which also includes MRPP1 and MRPP2 (ref. 10). MRPP1 and MRPP2 mRNA and protein levels were unchanged or increased in response to GTPP or CDDO (Fig. 2d), suggesting a rather specific downregulation of MRPP3 in the context of RNase P.

mtDNA-derived polycistronic pre-rRNA contains protein coding and ribosomal RNA elements flanked by tRNA genes. RNase P and RNase Z cleave pre-rRNA 5' and 3' of tRNAs, respectively, with 5' cleavage preceding 3' cleavage¹⁰. Consistent with reduced MRRP3 upon UPR^{mt},

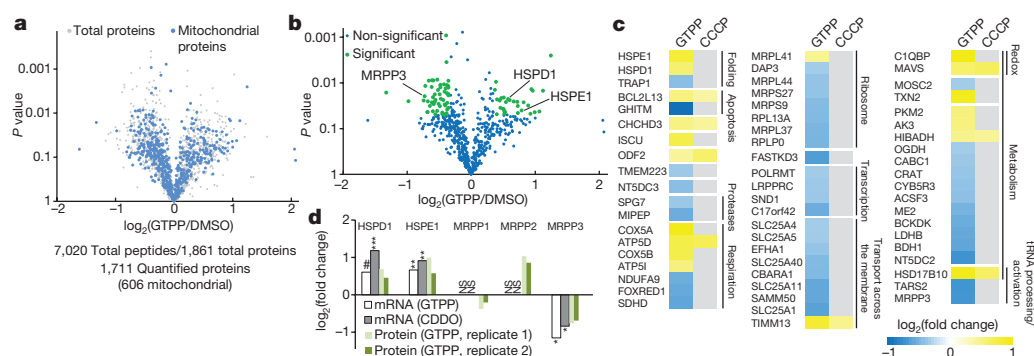


Figure 2 | Changes in the mitochondrial proteome upon UPR^{mt} induction. **a**, Volcano plot showing fold changes versus *P* values for total quantified and quantified mitochondrial proteins. **b**, Volcano plot showing fold changes versus *P* values for quantified mitochondrial proteins. Proteins significantly changing are indicated by green dots. **c**, Heatmap organized by gene ontology groups of mitochondrial protein level changes.

Proteins that did not change significantly are indicated in grey.
d, Histogram of protein (**b**) and/or mRNA (Fig. 1) abundance for chaperonin and mitochondrial RNase P subunits. Two-tailed *P* values * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, mean of $n = 3$ (RNA) or $n = 2$ (protein) biological replicates. NS, not significant. # $P = 0.06$.

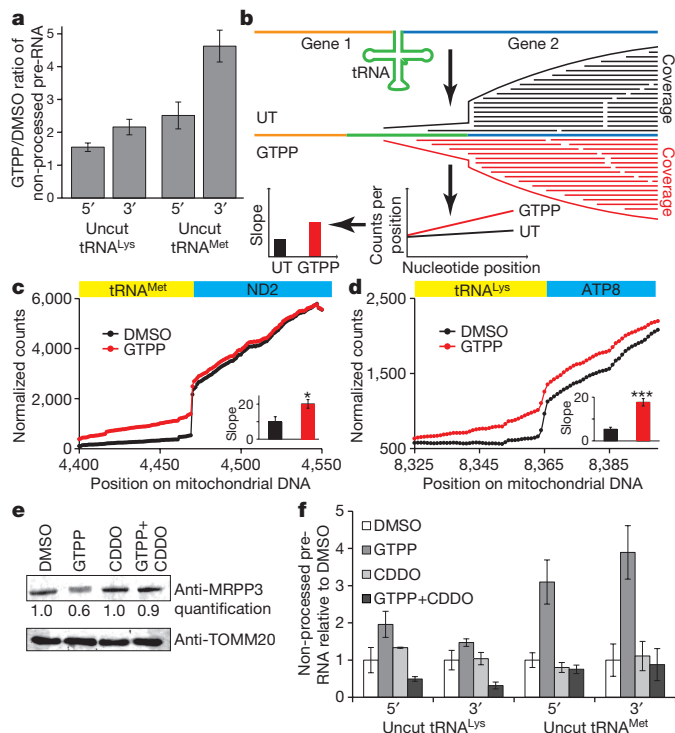


Figure 3 | Mitochondrial pre-RNA processing defects upon UPR^{mt}. **a**, qPCR of mitochondrial pre-RNA at tRNA^{Met} and tRNA^{Lys} RNaseP processing sites upon induction of UPR^{mt} with GTPP (6 h). Error bars, averages \pm s.d. ($n = 3$ biological replicates). **b**, RNA-seq for analysis of mitochondrial pre-RNA processing defects based on number of reads crossing the tRNA/mRNA gene junction. Slope of coverage in the tRNA gene adjacent to the cut site is used as a measure of processing. **c**, **d**, Normalized RNA-seq coverage across tRNA/mRNA gene borders for tRNA^{Met} (**c**) and tRNA^{Lys} (**d**) with average of slopes (\pm s.d.) as described in **b** indicated in the inset ($n = 3$ biological replicates, two-tailed P values * $P \leq 0.05$, *** $P \leq 0.001$). **e**, Quantitative western blot analysis of MRPP3 levels upon treatment of cells with dimethylsulfoxide (DMSO), GTPP, CDDO or GTPP + CDDO co-treatment for 6 h. **f**, Mitochondrial pre-RNA accumulation upon co-treatment with GTPP and CDDO (as in **a**). Data are average values \pm s.d. ($n = 3$ biological replicates). For gel source data, see Supplementary Fig. 1.

we observed a 1.5- to 4.5-fold increase in non-processed mitochondrial tRNA^{Lys} and tRNA^{Met} 6 h after GTPP treatment (Fig. 3a and Extended Data Fig. 6a), comparable to effects seen upon depletion of MRPP3 by siRNA (Extended Data Fig. 6b, c). To independently examine pre-RNA processing, we analysed coverage of pre-RNA cleavage sites in tRNA^{Met} and tRNA^{Lys} by RNA-seq. When pre-RNA processing is defective, sequence reads from the adjacent mRNA can extend into the tRNA, indicative of reduced processing as quantified via slopes of coverage (Fig. 3b). Indeed, upon UPR^{mt}, we observed increased slopes for sequence reads crossing *ATP8-tRNA^{Lys}* and *ND2-tRNA^{Met}* junctions (Fig. 3c, d). pre-RNA processing defects were absent with CCCP-dependent damage, suggesting a specific protein folding response (Extended Data Fig. 6d). While MRPP3 mRNA and protein levels are reduced upon treatment with GTPP, CDDO resulted in reduced MRPP3 mRNA levels without reduced MRPP3 protein levels (Figs 2d and 3e), suggesting LON-dependent MRPP3 degradation. Indeed, co-treatment with GTPP and CDDO resulted in no reduction in MRPP3 abundance (Fig. 3e) and, moreover, CDDO co-treatment rescued pre-RNA processing defects seen with GTPP alone (Fig. 3f). It is currently unclear how MRPP3 is made more susceptible to degradation in response to GTPP, but we conclude that this does not occur at the level of LON abundance, as LON is not increased upon GTPP treatment (Extended Data Fig. 6e, f).

We next examined whether loss of MRPP3 and defects in pre-RNA processing during UPR^{mt} could be overcome by its stable expression.

Previous studies have shown that alterations in the abundance of mitochondrial RNase P components can alter pre-RNA processing in unanticipated ways, making interpretation of effects of MRPP3 overexpression difficult¹⁹. Similarly, we found that elevated MRPP3 levels (~11-fold) altered steady-state processing efficiencies, with enhanced tRNA^{Met} processing and tRNA^{Lys} displaying enhanced 3' processing and decreased 5' processing (Extended Data Fig. 7a, b). While MRPP3 levels were still reduced upon GTPP treatment, consistent with LON activity not being limiting, residual MRPP3 remained approximately fivefold higher than in untreated cells (Extended Data Fig. 7a)¹⁹. Importantly, residual MRPP3 partly rescued tRNA^{Met} and tRNA^{Lys} processing (Extended Data Fig. 7c), consistent with the notion that loss of MRPP3 during UPR^{mt} contributes to pre-RNA processing defects.

UPR^{er} inhibits cytosolic translation through phosphorylation of eIF2 α and local degradation of mRNAs by IRE1 (ref. 11). The alterations in genes linked with mitochondrial protein synthesis (Figs 1 and 2) together with the finding that mitochondrial pre-RNA processing is deficient during UPR^{mt} led us to examine whether UPR^{mt} affects translation of mRNAs derived from mtDNA (Fig. 4a). Indeed, GTPP treatment (6 h) strongly inhibited ³⁵S-methionine incorporation into newly synthesized respiratory chain components in a concentration-dependent manner (Fig. 4b, c and Extended Data Fig. 8a) without affecting cytoplasmic translation rates (Extended Data Fig. 8b). To further validate the inhibitory effect of UPR^{mt} on mitochondrial translation, we used stable isotope labelling by amino acids in culture (SILAC) and mass spectrometry to quantify the ratio of newly synthesized (K8-Lys) to pre-existing (K0-Lys) protein for mitochondrially encoded proteins (Fig. 4d). Translational inhibition was confirmed for ND5, COI, ATP6 and ATP8 (Fig. 4d and Extended Data Figs 8c and 9a, b), with peptide coverage comparable to previous deep proteome studies in HeLa cells²⁰. Translational inhibition by GTPP, as well as pre-RNA processing, was largely recovered within 4 h of GTPP wash-out (Fig. 4e and Extended Data Fig. 10a, b), indicating that UPR^{mt} is rapidly reversible.

We find that acute mitochondrial folding stress promotes a multifaceted response involving (1) altered expression of nuclear genes, including mitochondrial chaperonins, to increase matrix protein folding capacity, (2) transcriptional repression and LON-dependent degradation of MRPP3 to reduce pre-RNA processing and (3) induction of rapid but reversible translational inhibition of mtDNA-encoded proteins, thereby reducing matrix folding load (Fig. 4f). As with pre-RNA processing (Extended Data Fig. 7b), cells overexpressing MRPP3 display altered translation of mtDNA-encoded proteins, with ND5, COI, ND2 and COIII showing decreased translation relative to control cells (Extended Data Fig. 10c), which complicates interpretation. However, residual MRPP3 post-GTPP treatment did not rescue bulk mitochondrial translation (Extended Data Fig. 10c). This could reflect sub-threshold levels of tRNA processing despite partial rescue (Extended Data Fig. 7) or redundancy in the UPR^{mt} pathway, thereby affecting other steps in the translation pathway (Fig. 4f), as is the case with UPR^{er} (ref. 11). Alternatively, because MRPP1-dependent tRNA methylation critical for tRNA maturation requires assembly with MRPP3 (ref. 21), MRPP3 overexpression may uncouple pre-RNA processing from tRNA methylation, resulting in translational defects despite the presence of MRPP3. While the TFB1M methyltransferase responsible for mitochondrial 12S rRNA methylation is reduced transcriptionally (Fig. 1h), its protein abundance is unchanged at 6 h post-GTPP (Extended Data Fig. 10d), indicating that defects in rRNA methylation do not underlie translational inhibition. Thus, further studies are required to understand the regulation of mitochondrial translation with and without mitochondrial stress. In keeping with the transient nature of stress responses¹³, our work has focused on acute effects of UPR^{mt}. Components linking mitochondrial protein misfolding to the nucleus remain to be identified, as ATFS-1 orthologues are lacking in mammals. Interestingly, the stress-inducible protein ATF3 (ref. 22), which contains a basic leucine zipper-like domain similar

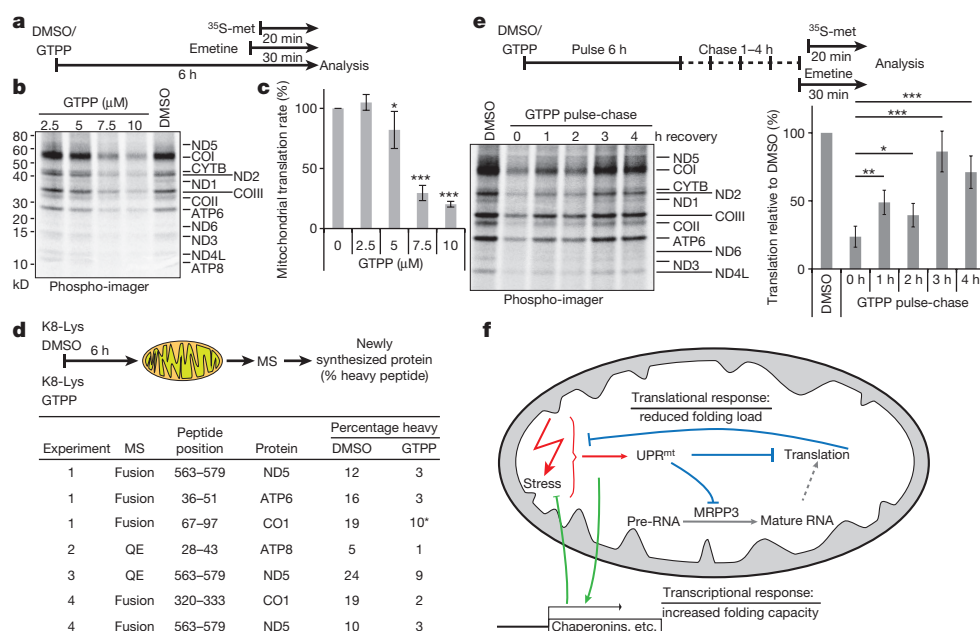


Figure 4 | UPR^{mt} halts mitochondrial translation. **a**, Experimental design. **b**, **c**, ³⁵S-methionine incorporation into newly translated proteins encoded by mtDNA with or without GTPP treatment was measured by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (**b**) and the gel quantified (**c**). Data are average values \pm s.d. across the 13 mitochondrially encoded proteins. Two-tailed *P* values **P* < 0.05, ****P* < 0.001. **d**, Heavy (K8)-Lys pulse assay for mitochondrial translation by SILAC mass spectrometry (top) was used to determine incorporation rates of K8-Lys into four mitochondrially encoded proteins, indicating relative translation rates (bottom); values were determined manually from the MS1 spectra (automatically retrieved values are shown in Extended Data Fig. 9c); *signal-to-noise ratios too low for accurate assessment. **e**, Reversibility of translation inhibition was measured using a GTPP pulse-chase assay wherein GTPP was washed out for 1–4 h (top). Translation rates are average values \pm s.d. across the 13 mitochondrially encoded proteins, two-tailed *P* values **P* < 0.05, ****P* < 0.001 and derived from SDS–PAGE of translation products (lower left) as described in **a**. **f**, Model of the cellular responses to UPR^{mt} induction. See text for details.

to ATF5-1, and which can function with CHOP, is also induced 1.5- to 4-fold by UPR^{mt} (Supplementary Table 1), suggesting a possible role in UPR^{mt} signalling. Prolonged UPR^{mt} and concomitant translational inhibition probably leads to confounding effects that would be detrimental to mitochondrial health, consistent with the application of GTPP to cancer therapeutics¹⁴. The transcriptional and proteomic data reported here provide a framework for the further elucidation of circuits that contribute to protein homeostasis within mitochondria, and for the development of approaches that can manipulate the response of cells to mitochondrial folding stress, as might occur in pathological conditions including cancer and neurodegenerative diseases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 November 2015; accepted 4 May 2016.

Published online 22 June 2016.

- Haynes, C. M., Petrova, K., Benedetti, C., Yang, Y. & Ron, D. ClpP mediates activation of a mitochondrial unfolded protein response in *C. elegans*. *Dev. Cell* **13**, 467–480 (2007).
- Haynes, C. M., Yang, Y., Blais, S. P., Neubert, T. A. & Ron, D. The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* **37**, 529–540 (2010).
- Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. Mitochondrial import efficiency of ATF5-1 regulates mitochondrial UPR activation. *Science* **337**, 587–590 (2012).
- Zhao, Q. *et al.* A mitochondrial specific stress response in mammalian cells. *EMBO J.* **21**, 4411–4419 (2002).
- Horibe, T. & Hoogenraad, N. J. The *chop* gene contains an element for the positive regulation of the mitochondrial unfolded protein response. *PLoS ONE* **2**, e835 (2007).
- Aldridge, J. E., Horibe, T. & Hoogenraad, N. J. Discovery of genes activated by the mitochondrial unfolded protein response (mtUPR) and cognate promoter elements. *PLoS ONE* **2**, e874 (2007).
- Haynes, C. M., Fiorese, C. J. & Lin, Y.-F. Evaluating and responding to mitochondrial dysfunction: the mitochondrial unfolded-protein response and beyond. *Trends Cell Biol.* **23**, 311–318 (2013).
- Kang, B. H. *et al.* Combinatorial drug design targeting multiple cancer signaling networks controlled by mitochondrial Hsp90. *J. Clin. Invest.* **119**, 454–464 (2009).
- Bernstein, S. H. *et al.* The mitochondrial ATP-dependent Lon protease: a novel target in lymphoma death mediated by the synthetic triterpenoid CDDO and its derivatives. *Blood* **119**, 3321–3329 (2012).
- Holzmann, J. *et al.* RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).
- Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).
- Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
- Lamech, L. T. & Haynes, C. M. The unpredictability of prolonged activation of stress response pathways. *J. Cell Biol.* **209**, 781–787 (2015).
- Siegelin, M. D. *et al.* Exploiting the mitochondrial unfolded protein response for cancer therapy in mice and human cells. *J. Clin. Invest.* **121**, 1349–1360 (2011).
- Guyton, K. Z., Xu, Q. & Holbrook, N. J. Induction of the mammalian stress response gene GADD153 by oxidative stress: role of AP-1 element. *Biochem. J.* **314**, 547–554 (1996).
- Zajac, M., Gomez, G., Benitez, J. & Martínez-Delgado, B. Molecular signature of response and potential pathways related to resistance to the HSP90 inhibitor, 17AAG, in breast cancer. *BMC Med. Genomics* **3**, 44 (2010).
- McAlister, G. C. *et al.* MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
- Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Lopez Sanchez, M. I. G. *et al.* RNA processing in human mitochondria. *Cell Cycle* **10**, 2904–2916 (2011).
- Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
- Reinhard, L., Sridhara, S. & Hällberg, B. M. Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res.* **43**, 5664–5672 (2015).
- Jiang, H. Y. *et al.* Activating transcription factor 3 is integral to the eukaryotic initiation factor 2 kinase stress response. *Mol. Cell. Biol.* **24**, 1365–1377 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. C. Altieri for GTPP, J. Paulo and S. P. Gygi for assistance with mass spectrometry, and L. Pontano and S. H. Sui for assistance with RNA-seq. This work was supported by National Institutes of Health grant R37NS083524 and Biogen, Inc. (J.W.H.), and an EMBO Fellowship (C.M.).

Author Contributions C.M. and J.W.H. were responsible for project conception, data analysis and manuscript preparation. All experiments were performed by C.M.

Author Information RNA sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE75411. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.H. (wade_harper@hms.harvard.edu).

Reviewer Information Nature thanks G. Shadel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Chemicals and antibodies. LysC (VWR catalogue number 100369-822), CDDO (Cayman Chemicals catalogue number 81035), emetine (Sigma catalogue number E2375), CCCP (Sigma catalogue number C2759), rotenone (Sigma catalogue number R8875), paraquat (Sigma catalogue number 36541), TFEA (Sigma catalogue number T27006), 3-Nitropropionic acid (Sigma catalogue number N5636), antimycin A (Sigma catalogue number A8674), myxothiazole (Sigma catalogue number T5580), potassium cyanide (Sigma catalogue number 60178), valinomycin (Sigma catalogue number V0627) and K8 lysine (Cambridge Isotopes). An original aliquot of GTPP was a gift from D. C. Altieri; a second aliquot was custom synthesized by Shanghai ChemPartner Co. Antibodies used were anti-MRPP3 (LSBio catalogue number LS-C332515, western blot: 1:500), anti-TOM20 (Santa Cruz catalogue number sc-11415, western blot: 1:500), anti-LON (Sigma catalogue number HPA002192, western blot 1:500), anti-ACTIN (Santa Cruz catalogue number sc69879, western blot 1:500), anti-TFB1M (Abcam, catalogue number 69871, western blot 1:400), anti-NDUFA9 (Abcam catalogue number ab14713, blue native 1:1,000), anti-SDHA (Abcam catalogue number ab14715, blue native 1:1,000) and anti-UQCRC2 (Abcam catalogue number ab14745, blue native 1:1,000).

Cell culture and assays for cytotoxicity, mitochondrial membrane potential and cellular ATP levels. HeLa cells were purchased from the American Type Culture Collection (ATCC) and not further authenticated. They were confirmed to be mycoplasma negative, and grown in RPMI medium supplemented with $1 \times$ glutamax (Invitrogen catalogue number 61870-127) and 10% fetal bovine serum. For all experiments, cells were treated with DMSO and $10 \mu\text{M}$ GTPP (or concentration as indicated) and/or $2.5 \mu\text{M}$ CDDO for 6 h. For CCK8 cytotoxicity assays, cells were plated in clear bottom 96-well plates, processed according to the manufacturer's instructions (CCK8 Dojindo CK04-05) and quantified on a VersaMax microplate reader (Molecular Devices). For mitochondrial membrane potential determination, cells were treated with JC-1 (Life Technologies catalogue number T3168) according to the manufacturer's instructions. Cells were harvested and analysed by fluorescence-activated cell sorting on a BD FACSCalibur. To assess cellular ATP levels, cells were plated on 96-well clear-bottom plates and treated with DMSO, GTPP or $100 \mu\text{M}$ antimycin A. ATP levels were measured with the Mitochondrial ToxGlo assay (Promega G8000) and analysed on a Molecular Devices SpectraMax M5 multi-mode plate reader.

Quantitative PCR and RNA sequencing. Total RNA was harvested using NucleoSpin RNA or NucleoSpin miRNA for analysis of pre-RNA processing (Macherey-Nagel catalogue numbers 740955 and 740971). RNA was quantified and equal amounts were reverse transcribed into complementary DNA (cDNA) using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems catalogue number 4368814). Quantitative PCR was performed using TaqMan Fast universal PCR Master Mix (Applied Biosystems catalogue number 4366072) or Fast SYBR Green Master Mix (Life Technologies catalogue number 4385612) using an Applied Biosystems 7500 Fast Real-time PCR machine with the following primers: tRNA^{Met} forward: AGTAAGGTCAGCTAAATAAG, tRNA^{Met} 5' upstream forward: GAATCGAACCCATCCCTGAG, tRNA^{Met} reverse: TAGTACGGGAAGGGTATAACC, tRNA^{Met} downstream reverse: GTGTGCCCTGCAAAGATGGTAG, tRNA^{Lys} forward: CACTGTAAAGCTAACTTAGC, tRNA^{Lys} 5' upstream forward: GAAATAGGCCCCGTATTTACC, tRNA^{Lys} reverse: TCACTGTAAAGAGGTGTTGG, tRNA^{Lys} 3' downstream reverse: GATGAGGAATAGTGTAAGGAG, GAPDH forward: ATGCCTCCTGCACCACCAAC, GAPDH reverse: GGGGCCATCCACAGTCTTCT, ND4 forward: CTTCCGAACCACTTATCC, ND4 reverse: GTATGCAATGACGCGATTTTAGG, or Life Technologies TaqMan probes for GAPDH (Hs02758991_g1), HSPD1 (Hs03044918_g1), HSP1 (Hs01654720_g1), MRPP3 (Hs00206448_m1), TRAP1 (Hs00212474_m1), DDIT3 (Hs00358796_g1), ATF4 (Hs00909569_g1) and BIP (Hs00607129_gH). For analysis of pre-RNA processing, data were normalized to tRNA levels obtained using internal forward and reverse tRNA primers. For analysis of integrated stress response activation, cells were treated with $10 \mu\text{M}$ GTPP, $10 \mu\text{g ml}^{-1}$ tunicamycin, $5 \mu\text{M}$ rotenone, 0.5 mM paraquat, 0.5 mM TFEA, 10 mM 3NP, $100 \mu\text{M}$ antimycin A, $3 \mu\text{M}$ myxothiazol, 1 mM KCN, $10 \mu\text{M}$ CCCP or $1 \mu\text{M}$ valinomycin for 6 h before purification and analysis of RNA levels by quantitative PCR.

For RNA sequencing, total RNA samples were submitted to the Harvard Bauer Core Facility for processing (ribosomal depletion with RiboZero), directional RNA-seq library preparation and 12 cycle amplification using LongAmp (New England BioLabs) and indexed primers (Integrated DNA Technologies), quality control and sequencing in one flow cell on a 75 bp paired-end NextSeq for transcriptome analysis or one lane on a 100 bp paired-end HiSeq to monitor mitochondrial RNA processing. For transcriptome analysis, reads were examined by FastQC and analysed by the tophat2 version 1.2 analysis pipeline by Harvard Medical School Research Computing against hg19, consisting of analysis by tophat, cufflinks, and cuffmerge. For analysis of mitochondrial pre-RNA processing, reads

were examined by FastQC, trimmed with cutadapt (for PHRED scores below five) and aligned to hg19 (augmented with transcript information from GRCh37.75) by STAR. Alignments were checked by FastQC and RNA-SeQC, and read counts of known genes detected by featureCounts.

To analyse the RNA-seq data set for pre-RNA processing, coverage data across a tRNA/mRNA region was normalized for reads within the protein-coding gene region across all six experimental conditions at every cut site. Slopes of the first ten nucleotides within the tRNA genes adjacent to the cut site were determined in Microsoft Excel (presented slopes had correlation values of $R \geq 0.9$) and calculated as an average of the mean with two-tailed P values.

Gene ontology enrichment analysis. Sets of genes of interest were uploaded and searched with the DAVID online tool (<http://david.abcc.ncifcrf.gov/home.jsp>) for enriched biological processes (GOTERM_BP_FAT). Functional annotation charts were saved and visualized with the enrichment map app (version 2.0.1)²³ in cytoscape (version 3.2.1, $P \leq 0.001$). Clusters were annotated according to their general functional with their overlapping biological process.

Promoter analysis. Three thousand bases upstream of the transcription start site of the transcripts encoding mitochondrial proteins and regulated by UPR^{mt} were extracted from ensemble. These promoter sequences were analysed by FIMO (Find Individual Motif Occurrences, version 4.11.1)²⁴. Motifs were provided as indicated and scanned on a provided DNA database with the listed promoter sequences. P values were set at 0.0001 and defined as the probability of random sequences of identical length achieving a similar or better score as the sequence provided.

RNAi experiments. Cells were grown on 12-well plates and RNAi was transfected with Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions. RNAi used was MRPP3 (Ambion, AM16708, ID 21858), and TRAP1 (DF/HCC DNA Resource Core IDs: HsSH00112394, HsSH00112407), EIF2AK1 (Dharmacon LQ-005007-00-0002), EIF2AK2 (Dharmacon LQ-003527-00-0002), EIF2AK3 (Dharmacon LQ-004883-00-0002) and EIF2AK4 (Dharmacon LQ-005314-00-0002).

Cell line generation. Human cDNA for MRPP3 was purchased from Sino Biological (HG14131-G) and transferred into a pHAGE lentiviral vector. Virus particles were produced in HEK293T cells after transfection with the lentiviral vector and helper vectors (VSVG, Tat1b, Mgpm2, CMV-Rev) and used to infect HeLa cells. Cells were selected in $1 \mu\text{g ml}^{-1}$ puromycin.

Mass spectrometry. For quantitative analysis of the mitochondrial proteome, HeLa cells were treated with $10 \mu\text{M}$ GTPP, CCCP, or DMSO for 6 h. Mitochondria were purified as previously described using Basic Protocol 1 (ref. 25). Briefly, cells were scraped into cold PBS, collected by centrifugation, resuspended in lysis buffer and sonicated. Crude mitochondria were acquired by differential centrifugation and purified mitochondria obtained by separation on a sucrose cushion. Similar amounts of mitochondria were obtained under the different treatments. Mitochondrial pellets were resuspended in lysis buffer (6 M GdnHCl, 75 mM NaCl, 50 mM Tris, pH 8.5, 1 mM PMSF, $1 \times$ OPT) and sonicated. Samples were reduced, alkylated with iodoacetamide and proteins were precipitated using chloroform/methanol. Protein pellets were resuspended in 8 M urea in 50 mM Tris, pH 8.8 and subsequently diluted with 50 mM Tris, pH 8.8 to a urea concentration of 2 M . Proteins were digested with LysC overnight at 37°C . Digestion reactions were stopped by addition of formic acid, dried and purified by C18 stage tip. Samples were taken up in 0.2 M HEPES, pH 8.5 buffer, quantified by micro BCA (Thermo Scientific catalogue number 23235) and labelled with TMT 6-plex reagents (Thermo Scientific) for 1 h at room temperature. Reactions were stopped by addition of 5% hydroxylamine for 15 min followed by addition of formic acid. Equal amounts of peptide samples were combined to a total of $10 \mu\text{g}$ and purified on a C18 stage tip. Dried peptides were resuspended in 5% acetonitrile/5% formic acid and analysed on an Orbitrap Fusion (Thermo Scientific) running a 2 h gradient from 6% to 30% acetonitrile using a multi-notch MS3-based method¹⁷ selecting ten MS2 fragment ions for analysis by MS3 (Orbitrap, AGC 5×10^4 , 60,000 resolution, maximum injection time 150 ms). Peptides were identified and quantified by a SEQUEST-based in-house tool (developed by the S. P. Gygi laboratory) using SEQUEST with a human UniProt database (as of 14 January 2014), and submitted to linear discriminant analysis to score peptides and proteins with protein and peptide FDR values of 2% (ref. 26). Proteins were collapsed to a protein-level FDR of 2%. Searches were run for LysC with a maximum of two missed cleavage sites and with carbamidomethylation of cysteine residues and TMT tags on lysine residues and N termini as static modifications, and methionine oxidation as variable modification. TMT-based quantitation was performed by TMT-reporter ion analysis for all identified proteins. MS3 spectra with a summed signal-to-noise ratio of <100 were excluded and the TMT channels normalized across all TMT channels (with resulting normalization factors between 1 and 1.252). For final analysis of quantified proteins, values were transferred and analysed in Microsoft Excel and the following cut-offs were applied: minimum number of two quantified

peptides, two-tailed $P \leq 0.05$, fold change $\geq \log_2 \pm 0.35$. Quantified proteins were determined as mitochondrial if they were found in MitoCarta²⁷, or a IMPI score of ≥ 0.85 (version Q1 2015, <http://www.mrc-mbu.cam.ac.uk/imp1>). For SILAC analysis, cell culture media was replaced with lysine-free media supplemented with K8 lysine and dialysed serum, and treated with DMSO or GTPP. After 6 h, cells were harvested and mitochondria purified, lysed and either processed as for TMT experiments (experiments 1–3), or run on a NuPAGE Novex 12% bis-tris gel (Life Technologies) and cut into five fractions, in-gel reduced, alkylated, and digested by LysC (experiment 4). Digested mitochondrial extracts and gel-extracted peptides were purified on C18 stage tips and analysed by LC-MS/MS on a Q Exactive or Orbitrap Fusion (Thermo Scientific) as indicated. Q Exactive analysis used a maximum injection time of 250 ms, an AGC target of 10^6 , resolution of 70,000 and automatic dynamic exclusion settings. For SILAC analysis on the Orbitrap Fusion, maximum injection times were set at 100 ms, AGC target at 2×10^5 , 120,000 resolution and a dynamic exclusion of 90 s. Experiments were processed with our in-house analysis tool and/or Maxquant (as indicated). For our in-house tool (Core), analysis was as above and quantification done by analysing peak heights for the heavy and light forms of a peptide. We performed MaxQuant analysis (version 1.5.2.8) with standard Orbitrap settings and LysC digestion mode with cysteine carbamidomethylation as static and methionine oxidation as variable modification against a UNIPROT library (as of 9 March 2015). The minimum ratio count of the protein quantification was set at 1. Results were exported into Microsoft Excel to calculate heavy-to-light ratios of peptides to determine the percentage of newly synthesized protein as a fraction of heavy peptide intensity versus total intensity. Results of the Maxquant/Core quantifications are shown in Extended Data Fig. 9c. Owing to the consistent difficulty of both analysis tools in determining heavy peptide intensities in the GTPP-treated samples, heavy and light peptide intensities were also manually determined from MS1 at the observed m/z values and retention times determined by the Maxquant/Core analyses (Fig. 4d and Extended Data Fig. 10).

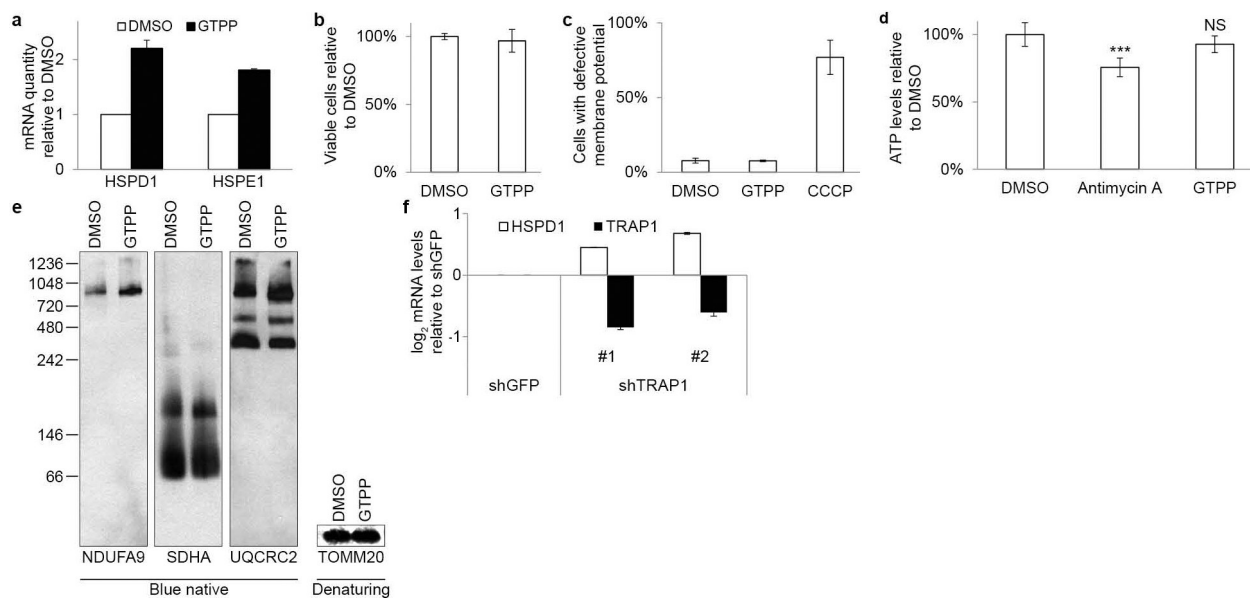
Mitochondrial translation assay. HeLa cells were grown on a 12-well plate and treated for 6 h with DMSO or different concentrations of GTPP as described. After 5.5 h, media was replaced with RPMI lacking methionine and containing 10% dialysed fetal bovine serum, GTPP or DMSO (at the original concentration), and $100 \mu\text{g ml}^{-1}$ emetine to block cytosolic translation. After 10 min, $100 \mu\text{C ml}^{-1}$

EasyTag ³⁵S-methionine (Perkin Elmer catalogue number NEG709A500UC) was added and incubated for another 20 min, totalling 6 h of GTPP treatment. Cells were washed with PBS and harvested in $1 \times$ NuPAGE LDS sample buffer (Life Technologies) containing 25 mM DTT. Samples were boiled and analysed on a NuPAGE Novex 12% bis-tris gel (Life Technologies). Gels were stained using InstantBlue (Expedeon), dried onto Whatman paper and visualized on a Bio-Rad Personal Molecular Imager for newly synthesized and radioactive proteins. An image was taken of the InstantBlue stained gel to confirm equal loading. These experiments were performed three independent times. For pulse-chase analysis, the same protocol was used with washes as indicated. This experiment was performed twice independently.

Blue native. Crude mitochondria were obtained as above and lysed in 1% digitonin, followed by separation on 4–16% BN-PAGE as previously described²⁸. Proteins were transferred onto polyvinylidene difluoride membranes and detected using antibodies as indicated. A small aliquot was also analysed by standard western blot to confirm equal loading.

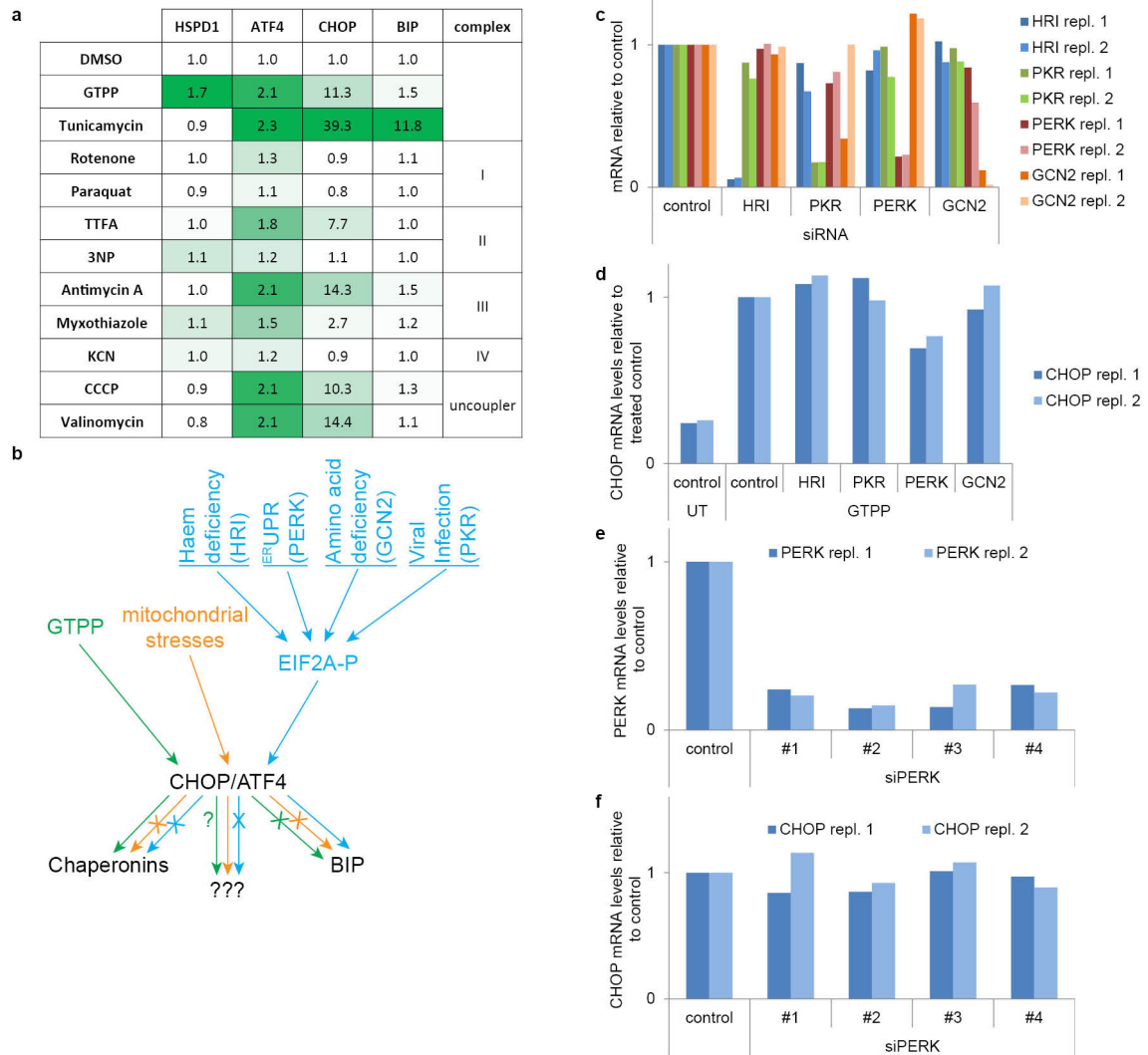
Data reporting and statistics. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. All quantitative experiments are presented as means \pm s.d. of at least two independent biological experiments (as indicated) and were analysed by a two-tailed Student's t -test (considered significant for $P \leq 0.05$).

23. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).
24. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
25. Bozidis, P., Williamson, C. D. & Colberg-Poley, A. M. in *Current Protocols in Cell Biology* (eds Bonifacino, J. S. et al.) Ch. 3, Unit 3.27 (Wiley, 2007).
26. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
27. Pagliarini, D. J. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
28. McKenzie, M., Lazarou, M., Thorburn, D. R. & Ryan, M. T. Mitochondrial respiratory chain supercomplexes are destabilized in Barth syndrome patients. *J. Mol. Biol.* **361**, 462–469 (2006).



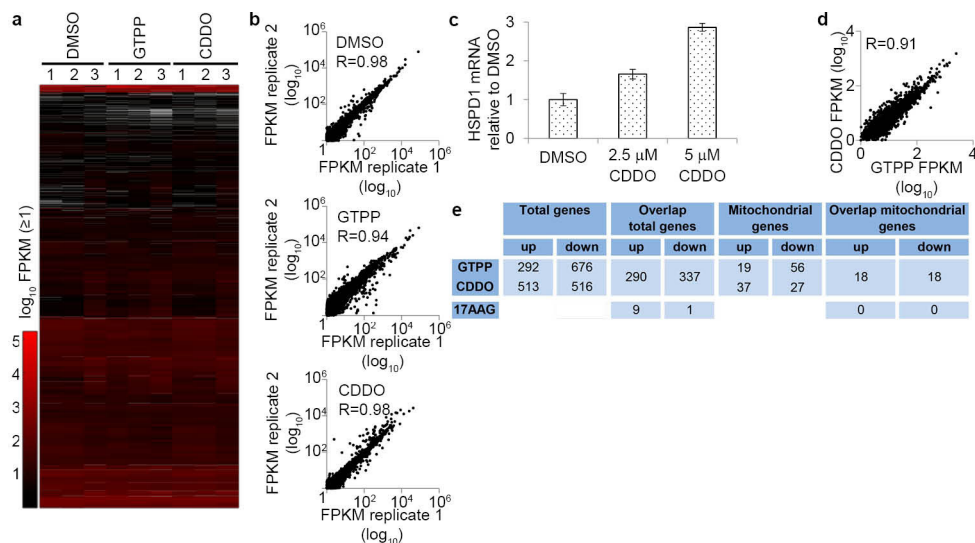
Extended Data Figure 1 | Mitochondrial HSP90 inhibition induces UPR^{mt}. **a**, Quantitative PCR monitoring chaperonin (*HSPD1* and *HSPD1*) mRNA levels upon treatment of cells with GTPP. Shown are means of levels relative to untreated \pm s.d. ($n = 3$ biological replicates). **b**, Measurement of cell viability upon GTPP treatment with CCK8. Shown are means of levels relative to untreated \pm s.d. ($n = 5$ biological replicates). **c**, Measurement of mitochondrial membrane potential upon GTPP or CCCP (mitochondrial membrane potential uncoupler) treatment, measured with JC-1 and analysed on a BD FACSCalibur. Shown are means of levels relative to untreated \pm s.d. ($n = 3$ biological replicates).

d, Measurement of cellular ATP levels upon GTPP or antimycin A (electron transport chain inhibitor) treatment. Shown are means of levels relative to untreated \pm s.d. ($n = 4$ biological replicates) and two-tailed P values *** $P \leq 0.001$; NS, not significant. **e**, Blue native gel analysis of mitochondrial respiratory chain complexes upon 6 h treatments of DMSO or GTPP. **f**, Changes in chaperonin and *TRAP1* mRNA levels upon knockdown with shRNA targeting *GFP* or *TRAP1* mRNA. Shown are means of log₂ fold changes relative to control and s.d. ($n = 3$ biological replicates).



Extended Data Figure 2 | UPR^{mt} signals distinctly from the integrated stress response. **a**, Table with summarized results of data shown in Fig. 1a–c. Induced genes are labelled green and compounds are clustered into their molecular function. GTPP induces UPR^{mt} and tunicamycin ERUPR. All other compounds affect mitochondrial respiration and/or the mitochondrial membrane potential. **b**, Schematic showing how different stresses signal through the integrated stress response pathway based on the results shown in **b** and Fig. 1a–c. **c**, Quantitative PCR to assess the mRNA

knockdown of the four EIF2A kinases by siRNA smart pools in biological duplicate; repl., replicate **d**, Quantitative PCR monitoring *CHOP* mRNA levels in untreated or GTPP-treated cells with or without knockdown of the EIF2A kinases as in **c**. **e**, Quantitative PCR to monitor *PERK* mRNA levels upon *PERK* knockdown with individual siRNAs in biological duplicate. **f**, Quantitative PCR monitoring *CHOP* mRNA levels in GTPP-treated cells with or without knockdown of *PERK* by individual siRNAs in biological duplicate.

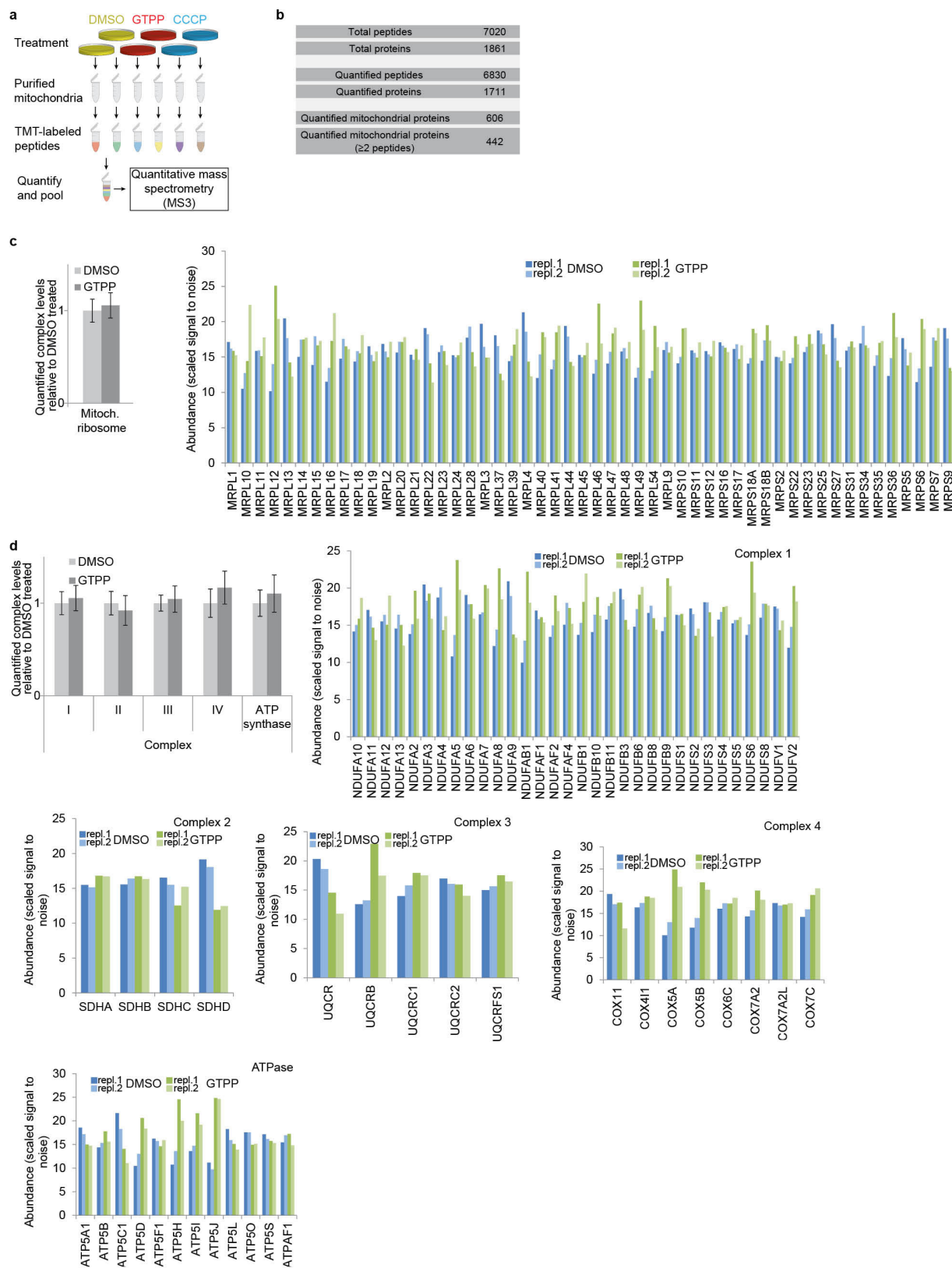


Extended Data Figure 3 | Global analysis of transcriptional responses to UPR^{mt} induction. **a**, Heatmap of measured transcript abundances of cells treated with DMSO, 10 μ M GTPP or 2.5 μ M CDDO for 6 h ($n = 3$ biological replicates). Values not passing the cuffdiff threshold of FPKM abundance and read number were excluded (white). **b**, Correlation of replicates for DMSO-, GTPP- and CDDO-treated samples with R values depicting correlation value; \log_{10} -transformed FPKM values (≥ 0) are plotted. **c**, Quantitative PCR monitoring induction of ^{mt}UPR

by measuring chaperonin mRNA levels upon treatment with DMSO or CDDO. Shown are means of levels relative to DMSO-treated \pm s.d. ($n = 3$ biological replicates). **d**, Correlation between the abundance of transcripts significantly altered in GTPP- versus CDDO-treated cells (Fig. 1c, combined panel). **e**, Table representing changed transcripts upon GTPP or CDDO treatment (Fig. 1c) compared with the number of transcripts changed upon 17AAG previously reported¹⁶.

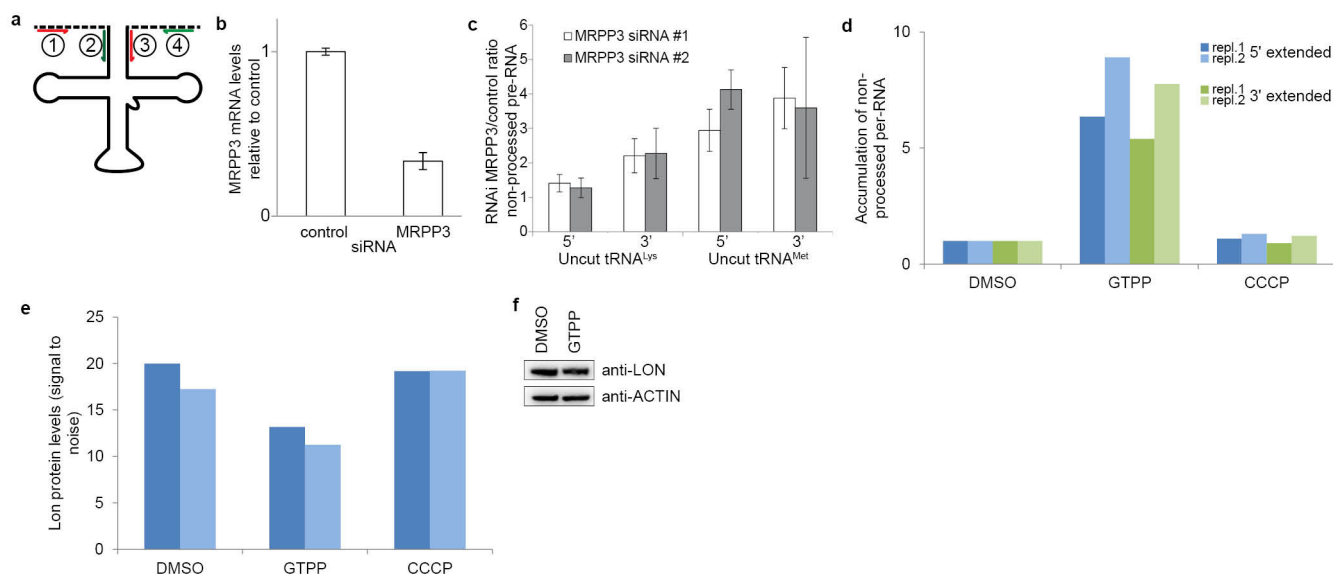
Gene	CHOP (D)DVNTGCAHYN	MURE1 AGAATBGCT	MURE2 GYACBCSAG	ATF4 VTKNCDHMR
ABCB10				
ABCD3				
ALDH1L2				
C22ORF32				
COX8A				
EARS2				
ECSIT				
FAM36A				
FASTKD2				
GARS				
GFER				
GPT2				
HSPD1				
HSPE1				
IDI1				
IMMP2L				
IREB2				
LDHAL6B				
MARS				
MRPL18				
MRPL22				
MRPP3				
MRPS31				
MTHFD2				
NDUFA11				
OAT				
PDK3				
PDK4				
PRELID2				
PTGS2				
SARS				
SLC22A4				
SLC25A12				
SLC25A40				
SOD1				
TFB1M				

Extended Data Figure 4 | Promoter analysis of UPR^{mt}-induced transcripts encoding mitochondrial proteins. Analysis of UPR^{mt}-induced (GTPP and CDDO) transcripts encoding mitochondrial proteins for the occurrence of CHOP, MURE1, MURE2, or ATF4 promoter elements. We used FIMA version 4.11.1 with the consensus sequences shown. Cells marked in green represent the presence of the consensus sequence in the gene shown.



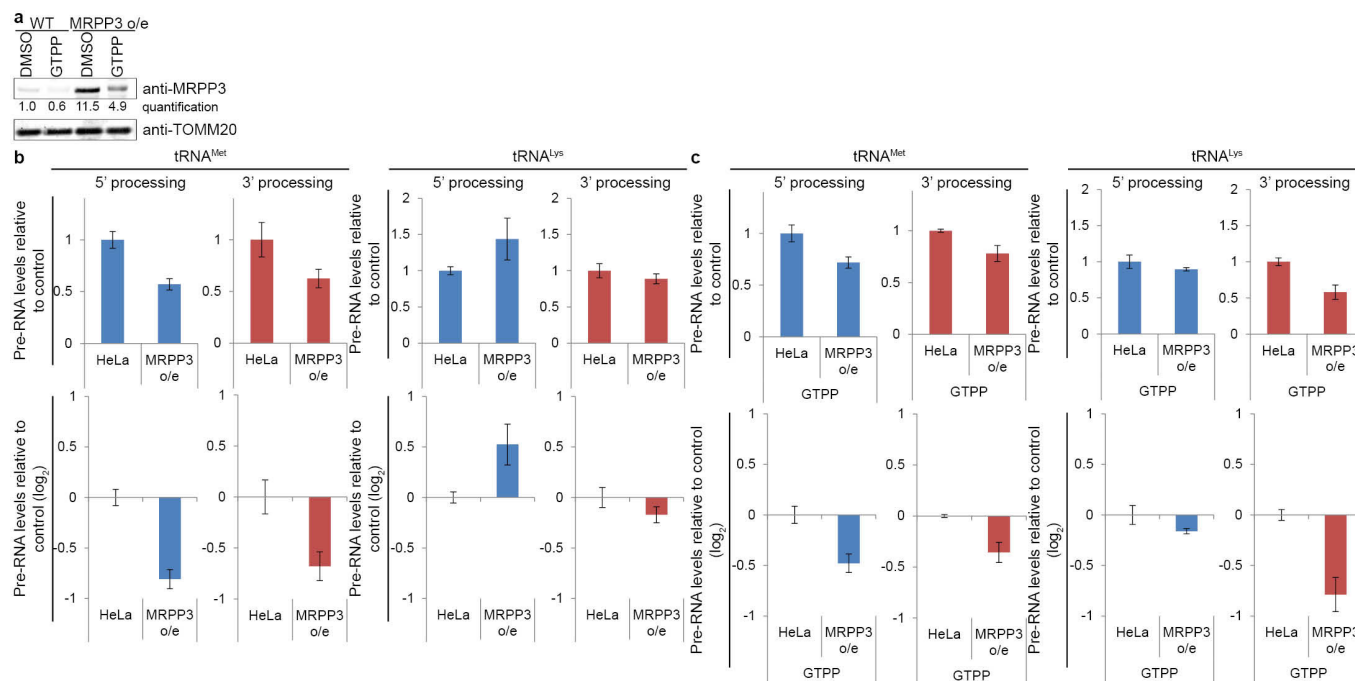
Extended Data Figure 5 | Changes in the mitochondrial proteome upon UPR^{mt} induction. **a**, Assay design. **b**, Summary of proteomic data. **c**, Analysis of changes in the average abundance of mitochondrial ribosome (left) or for individual ribosomal subunits (right). Values are mean values \pm s.d. of scaled signal to noise values (that is, relative abundance) derived from the quantitative proteomics (Fig. 2) for identified mitochondrial ribosomal subunits (right, $n = 2$ biological

replicates) and the average of all these values \pm s.d. (left); repl., replicate. **d**, Analysis of the abundance of the different mitochondrial electron transport chain complexes and ATP synthase. Values are derived from quantitative proteomics (Fig. 2) and shown as mean values \pm s.d. across all quantified subunits (top left) or separately per subunit for the different complexes ($n = 2$ biological replicates). All data depict scaled signal to noise values (that is, relative abundance).



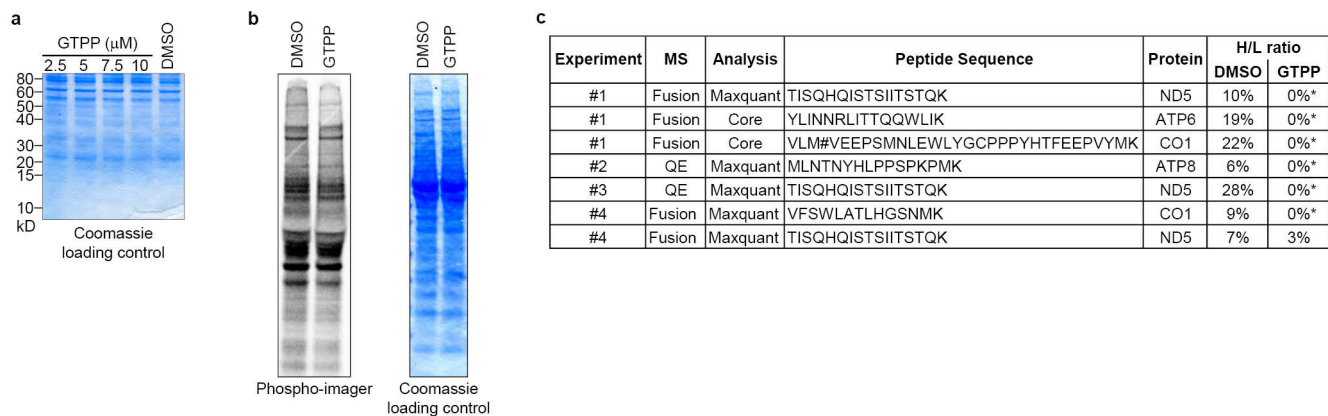
Extended Data Figure 6 | Mitochondrial pre-RNA processing defects upon UPR^{mt}. **a**, Primer design for monitoring pre-RNA processing. Primer pairs 1 and 3, and 2 and 4, will only produce PCR products for uncleaved mitochondrial pre-RNAs and allow quantitation of non-processed pre-RNAs. Primer pair 2 and 3 will monitor total levels for normalization. **b**, Quantitative PCR of *MRPP3* mRNA levels upon knockdown with siRNA targeting a scrambled sequence or *MRPP3*. Shown are averages \pm s.d. ($n = 3$ biological replicates). **c**, qPCR of

mitochondrial pre-RNA at *tRNA^{Met}* and *tRNA^{Lys}* RNaseP processing sites upon depletion of *MRPP3* by RNAi. Error bars, \pm s.d. ($n = 3$ biological replicates). **d**, Quantitative PCR monitoring levels of non-processed pre-RNA upon treatment of cells with GTPP or the uncoupler CCCP in biological duplicate; repl., replicate. **e**, LON protein levels as determined by quantitative proteomics (Fig. 2) in biological duplicate. Shown are scaled signal to noise values observed (that is, relative abundance). **f**, Western blot analysis of LON levels upon control or 10 μ M GTPP treatment (6 h).



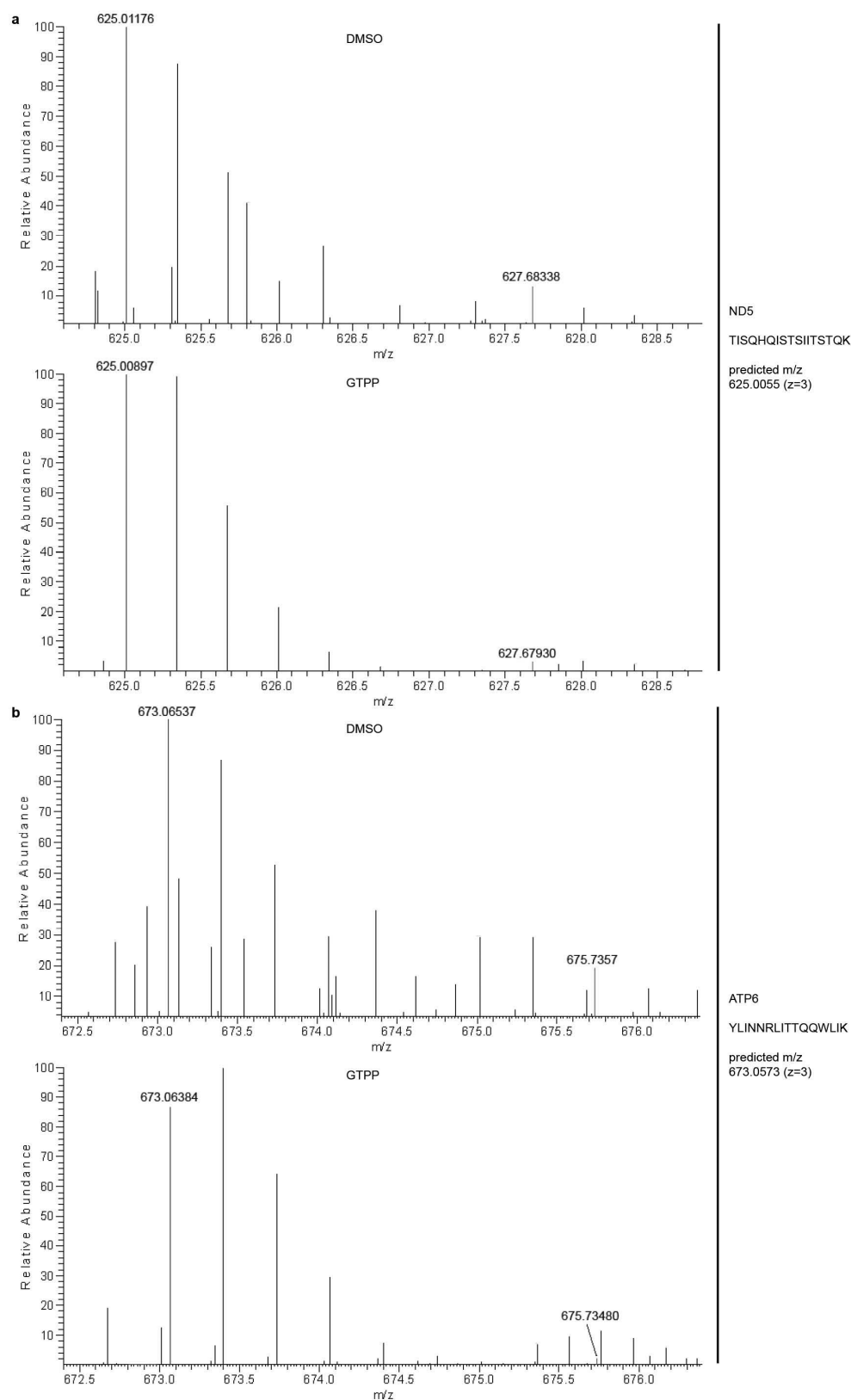
Extended Data Figure 7 | Rescue of UPR^{mt}-induced mitochondrial pre-RNA processing by MRPP3 overexpression. **a**, Western Blot analysis of MRPP3 levels upon DMSO or GTPP treatment in the context of wild-type or MRPP3-overexpressing (o/e) cells. **b**, Quantitative PCR analysis of non-processed mitochondrial pre-RNA levels at the tRNA^{Met} and tRNA^{Lys}

cut sites in wild-type cells or cells overexpressing MRPP3. Shown are mean values \pm s.d. ($n = 3$ biological replicates). **c**, Quantitative PCR analysis of non-processed mitochondrial pre-RNA levels at the tRNA^{Met} and tRNA^{Lys} cut sites in wild-type cells or cells overexpressing MRPP3 upon GTPP treatment. Shown are mean values \pm s.d. ($n = 3$ biological replicates).

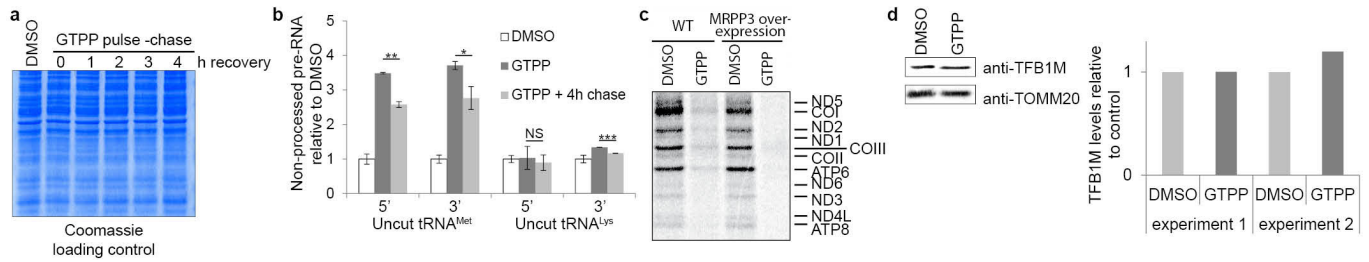


Extended Data Figure 8 | Mitochondrial translation defects upon UPR^{mt}. **a**, Coomassie gel staining as a loading control of the same experiment as in Fig. 4b. **b**, Analysis of cytosolic translation upon treatment with DMSO or GTPP with the same experimental procedure as in Fig. 4a without the addition of emetine. Newly synthesized proteins were monitored by phospho-imager (left) with Coomassie staining of the same gel as loading control (right). **c**, Table of experiment number, mass

spectrometer used, analysis method, peptides sequence, protein encoded and heavy-to-light ratios (H/L) used to determine protein synthesis rates in Fig. 4d. Fusion and QE are Orbitrap Fusion or Q Exactive (Thermo Scientific), respectively; Core depicts in-house mass spectrometry analysis pipeline; #, oxidative modification on methionine; *, could not be determined by Core/Maxquant.



Extended Data Figure 9 | Proteomic determination of mitochondrial translation upon UPR^{mt}. SILAC spectra for the data shown in Fig. 4d and Extended Data Fig. 9c for experiment 1.



Extended Data Figure 10 | Reversibility of UPR^{mt}-induced mitochondrial pre-RNA processing and translation defects.

a, Coomassie gel staining as a loading control of the same experiment as in Fig. 4e. **b**, Mitochondrial pre-RNA processing was measured by qPCR in cells subjected to GTPP pulse-chase wash-out for 1–4 h. Data are averages of fold changes versus untreated \pm s.d., two-tailed P values $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $n = 3$ biological replicates. **c**, Analysis

of mitochondrial translation in wild-type or MRPP3 overexpressing cells with or without GTPP treatment. Newly synthesized proteins were labelled with ^{35}S and analysed by phospho-imager. **d**, Immunoblot of TFB1M expression with or without 6 h GTPP treatment (left). Quantification of control normalized TFB1M levels from immunoblots of two independent experiments.

H4K20me0 marks post-replicative chromatin and recruits the TONSL–MMS22L DNA repair complex

Giulia Saredi^{1*}, Hongda Huang^{2*}, Colin M. Hammond¹, Constance Alabert¹, Simon Bekker-Jensen³, Ignasi Forne⁴, Nazaret Reverón-Gómez¹, Benjamin M. Foster⁵, Lucie Mlejnkova⁶, Till Bartke⁵, Petr Cejka⁶, Niels Mailand³, Axel Imhof⁴, Dinshaw J. Patel² & Anja Groth¹

After DNA replication, chromosomal processes including DNA repair and transcription take place in the context of sister chromatids. While cell cycle regulation can guide these processes globally, mechanisms to distinguish pre- and post-replicative states locally remain unknown. Here we reveal that new histones incorporated during DNA replication provide a signature of post-replicative chromatin, read by the human TONSL–MMS22L^{1–4} homologous recombination complex. We identify the TONSL ankyrin repeat domain (ARD) as a reader of histone H4 tails unmethylated at K20 (H4K20me0), which are specific to new histones incorporated during DNA replication and mark post-replicative chromatin until the G2/M phase of the cell cycle. Accordingly, TONSL–MMS22L binds new histones H3–H4 both before and after incorporation into nucleosomes, remaining on replicated chromatin until late G2/M. H4K20me0 recognition is required for TONSL–MMS22L binding to chromatin and accumulation at challenged replication forks and DNA lesions. Consequently, TONSL ARD mutants are toxic, compromising genome stability, cell viability and resistance to replication stress. Together, these data reveal a histone-reader-based mechanism for recognizing the post-replicative state, offering a new angle to understand DNA repair with the potential for targeted cancer therapy.

The TONSL–MMS22L complex is an obligate heterodimer required for replication fork stability and repair of replication-associated DNA damage by aiding RAD51 loading^{1–4}. TONSL–MMS22L associates with soluble non-nucleosomal histones H3–H4 (refs 1, 5), the histone chaperone ASF1 (refs 1–4) and MCM2/4/6/7 (refs 1–5) in a manner that depends on the TONSL ARD¹. We have found that histones H3–H4 bridge the interactions between TONSL–MMS22L and ASF1 (ref. 1), between ASF1 and MCM2 (refs 6, 7), and between TONSL–MMS22L and MCM2 (Extended Data Fig. 1a), suggesting simultaneous binding of these proteins to histones H3–H4 in a large pre-deposition complex. In addition, TONSL–MMS22L interacts with nucleosomal histones in chromatin (Extended Data Fig. 1b). This suggests that TONSL–MMS22L functions as an H3–H4 histone chaperone while also acting as a histone reader in chromatin. Consistent with a chaperone function, TONSL was recently shown to have histone chaperone activity *in vitro*⁵. We therefore set out to explore the mechanism of action of TONSL–MMS22L by a structure–function approach.

Full-length TONSL and the ARD alone bound directly to histones H3–H4 but not H2A–H2B *in vitro* (Extended Data Fig. 1c–f). As our attempts to crystallize the ARD with H3–H4 were not successful, we linked the ARD to the MCM2 histone-binding domain (HBD), because

a similar design previously enabled us to solve the structure of an H3–H4 dimer in complex with MCM2 and ASF1 (ref. 7). We obtained crystals of covalently linked MCM2 HBD–G4–TONSL ARD in complex with H3 (57–135) and H4 that diffracted to 2.43 Å resolution, and solved the structure by molecular replacement on the basis of our structure of MCM2 HBD in complex with an H3–H4 tetramer⁷ (Fig. 1a, b; for X-ray statistics, see Extended Data Table 1). The structure shows a pair of MCM2 HBDs wrapped around the lateral surface of the H3–H4 tetramer, similar to the MCM2–HBD–H3–H4 complex alone^{7,8}, while two TONSL ARDs interact with each of the H4 tails (Fig. 1a, b). The G4-linker along with flanking residues formed a 19-residue-long disordered segment that could reach a distance of up to 70 Å. The distance from the observed C terminus of MCM2 HBD to the observed N terminus of TONSL ARD is only 10 Å, indicating that the covalent linkage within the MCM2–HBD–G4–TONSL–ARD cassette does not affect the structural integrity of the complex. TONSL ARD forms no intermolecular interactions with the MCM2 HBD, consistent with H3–H4 bridging the interaction of TONSL and MCM2 in cells (Extended Data Fig. 1a), and it shows only minimal contacts with the core of the H3–H4 tetramer (Fig. 1a, b). However, the TONSL ARD forms extensive contacts with a segment of the H4 tail (Fig. 1b, c and Extended Data Fig. 1g) and, consistently, it binds the histone H4 tail, but not the H3 tail, *in vitro* (Extended Data Fig. 2a). In addition to defining TONSL binding to soluble histone H3–H4 in complex with MCM2 (Fig. 1a, b), this binding mode is also compatible with TONSL binding histone H3–H4 dimers in a co-chaperone complex with MCM2 and ASF1 (ref. 7) as well as recognizing H4 tails in a nucleosome (see models in Extended Data Fig. 2b, c).

The TONSL ARD consists of four ankyrin repeats, three of which adopt the canonical fold (ANK1–3), while the remaining one is an atypical and capping repeat (ANK4) (Extended Data Fig. 1g). The TONSL ARD uses its elongated concave surface to form extensive contacts with the H4 tail in an extended β -strand-like conformation (Fig. 1c and Extended Data Fig. 1g). Notably, 15 out of 18 residues that constitute the H4-tail-binding surface of TONSL ARD are highly conserved (Extended Data Fig. 2d). The TONSL ARD targets the H4 tail, spanning residues Lys12 to Arg23 (Fig. 1c–g and Extended Data Fig. 3a, b) with three consecutive binding channels accommodating Arg17, His18 and Lys20 (Fig. 1d). These H4 residues are part of a basic region, which can interact with the acidic patch on neighbour nucleosomes⁹ in compact chromatin. H4 Arg17 forms two hydrogen bonds with ARD Asn571 and stacks with Tyr572 and Cys608 (Fig. 1c, e), while H4 His18 penetrates into a pocket lined by four strictly conserved residues (Trp563, Glu568, Asn571 and Asp604) (Fig. 1c, f). Substitution of H4 His18 with the larger Trp residue (mutant H18W) disrupts

¹Biotech Research and Innovation Centre (BRIC) and Centre for Epigenetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark. ²Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen DK-2200, Denmark. ⁴Department of Molecular Biology, Biomedical Center and Center for Integrated Protein Science Munich, Ludwig-Maximilians University, 80336 Munich, Germany. ⁵MRC Clinical Sciences Centre (CSC) and Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK. ⁶Institute of Molecular Cancer Research, University of Zurich, Zurich CH-8057, Switzerland.

*These authors contributed equally to this work.

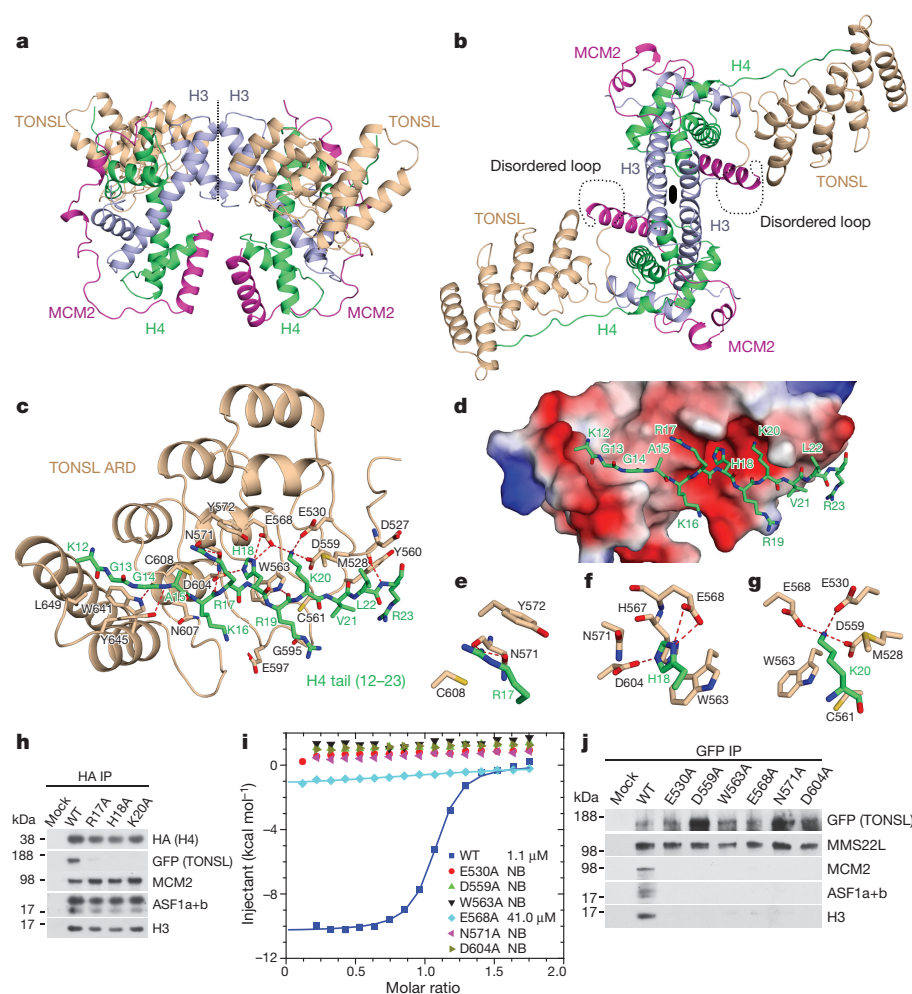


Figure 1 | TONSL ARD interacts with the histone H4 tail. **a, b**, Two different representative views of the overall structure of the TONSL-ARD-MCM2-HBD-H3-H4 tetramer complex. **c**, Intermolecular interactions between TONSL ARD and the H4 tail. **d**, The electrostatic potential surface of ARD showing the acidic concave surface binding site for the H4 tail. **e–g**, Highlights of the intermolecular interactions of H4 Arg17, His18 and Lys20 with ARD. **h**, Immunoprecipitation (IP) of soluble haemagglutinin (HA)-SNAP-H4 wild type (WT) or mutant transfected into green fluorescent protein (GFP)-TONSL U-2-OS cells. **i**, ITC of TONSL ARD wild type and mutants with H4 tail peptide. **j**, Immunoprecipitation of soluble GFP-TONSL wild type or mutant. **h, j**, Data are representative of three independent experiments. For protein inputs, see Extended Data Fig. 9b, c; for gel source data, see Supplementary Fig. 1.

binding with ARD (Fig. 2a), underscoring the importance of fitting His18 in the pocket. The H4 Lys20 residue is bound within an acidic surface channel on ARD (Fig. 1c, d). The side chain of H4 Lys20 interacts with Met528 and contacts the edge of Trp563 of ARD, while the main-chain atoms of H4 Lys20 packs against Cys561 of ARD (Fig. 1g). The N ϵ atom of H4 Lys20 forms three strong hydrogen bonds (distance <3 Å) with the side chains of strictly conserved residues Glu530, Asp559 and Glu568 of ARD, which engage H4 Lys20 in a triangular arrangement (Fig. 1g). Consistent with the structural data, histone

H4 mutations R17A, H18A and K20A disrupted binding to TONSL in cells (Fig. 1h). Likewise, mutation of six conserved TONSL residues lining the H4 Arg17, His18 and Lys20 binding channels disrupted binding to H4 peptides and recombinant histone H3–H4 (Fig. 1i and Extended Data Fig. 3c). *In vivo*, these mutants abrogated binding to soluble histone H3–H4 and, consequentially, also association with ASF1a and ASF1b and MCM2 without affecting MMS22L binding to TONSL^{1,2} (Fig. 1j and Extended Data Fig. 3d). These mutations did not affect ARD structure, as indicated by circular dichroism (Extended Data Fig. 3e).

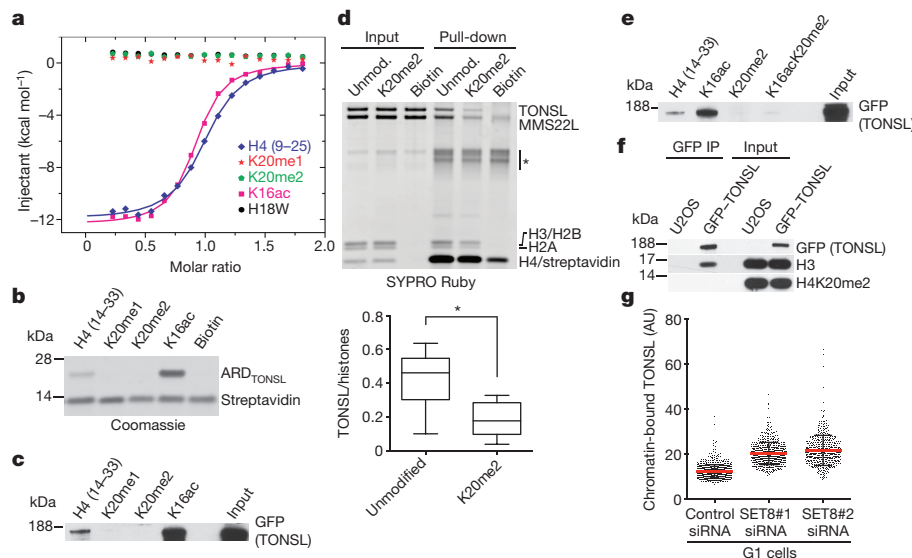


Figure 2 | TONSL ARD recognizes unmodified H4K20. **a**, ITC of TONSL ARD binding to H4 tail peptides. **b, c**, Pull-down of recombinant TONSL ARD (b) or GFP-TONSL from extracts (c) with biotinylated H4 tail peptides. **d**, Top, pull-down of recombinant TONSL-MMS22L with biotinylated recombinant mononucleosomes. Asterisk indicates unspecific band. Unmod., unmodified. Bottom, TONSL binding quantified relative to histones. Unpaired *t*-test: **P* < 0.05; mean *n* = 6; whiskers, outliers. **e**, As in c. **f**, Immunoprecipitation of GFP-TONSL from solubilized chromatin. **g**, TONSL chromatin binding in pre-extracted SET8-depleted G1 cells. AU, arbitrary units. Error bars indicate standard deviation (s.d.); *n* = 747 (short interfering RNA (siRNA) control), 579 (SET8#1) and 485 (SET8#2). Data are representative of four (b), three (c) and two (e–g) independent experiments. For peptide inputs, see Extended Data Fig. 9d.

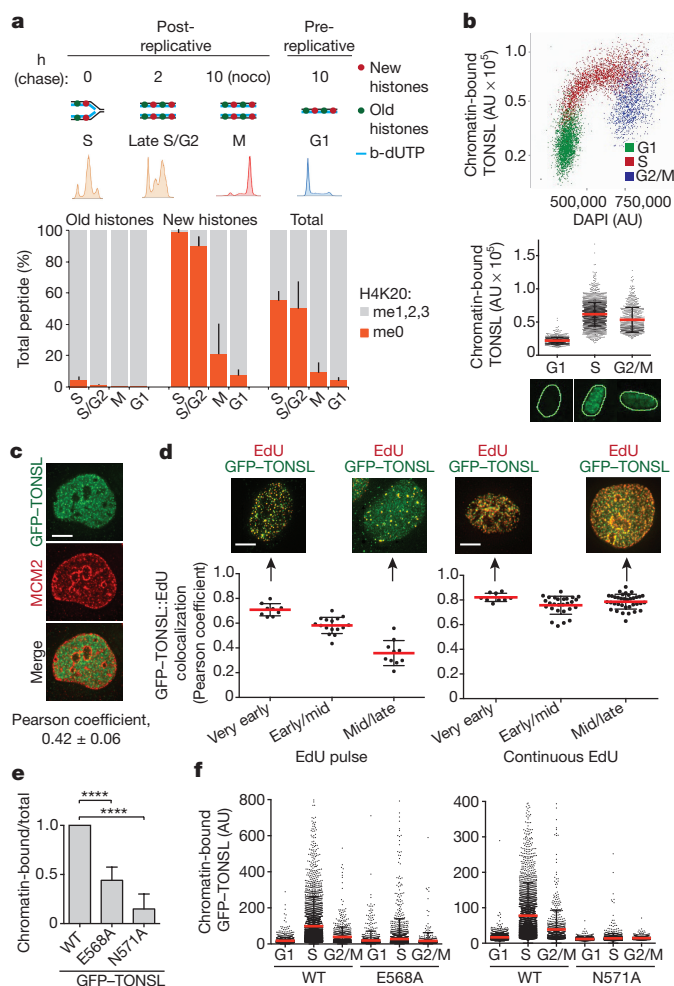


Figure 3 | H4K20me0 is a signature of post-replicative chromatin, read by TONSL ARD. **a**, H4K20me on new and old histones analysed by stable isotope labelling with amino acids in cell culture (SILAC)-based mass spectrometry of chromatin pulse-labelled with biotin-dUTP (b-dUTP) and isolated by NCC (data from ref. 18). noco, nocodazole. Error bars indicate s.d.; $n = 9$ (S), 3 (S/G2, M), 5 (G1); M (old histones) shows the mean of $n = 2$; see also Extended Data Fig. 6a, b. **b**, TONSL chromatin binding in pre-extracted TIG3 fibroblasts shown as a function of 4',6-diamidino-2-phenylindole (DAPI) intensity or cell cycle stage. Error bars indicate s.d.; $n = 886$ (G1), 2194 (S) and 756 (G2); representative cells are shown. **c**, **d**, Colocalization analysis of chromatin-bound GFP-TONSL with MCM2 (**c**) and EdU (**d**). **d**, Cells were pulsed with EdU (left) or released into S phase in the presence of EdU (right) and analysed by deconvolution microscopy. Error bars indicate s.d.; **c**, $n = 13$; **d**, from left, $n = 9, 16, 10, 9, 27, 36$; representative cells are shown. Scale bar, 5 μm . **e**, Chromatin binding of GFP-TONSL analysed by cellular fractionation, quantified relative to total GFP-TONSL and normalized to wild type (WT). Error bars indicate s.d.; $n = 5$ (wild type/N571A); $n = 3$ (E568A). Unpaired t -test: **** $P < 0.0001$. **f**, Chromatin binding of GFP-TONSL analysed as in **b**. Error bars indicate s.d.; from left, $n = 1,302, 3,567, 750, 1,311, 3,850, 838, 1,495, 3,221, 832, 1,695, 2,729, 877$. Data are representative of two (**b–d**, **f**) independent experiments.

Together, this defines TONSL ARD as a recognition module for histone H4 tails, distinct from the GLP/G9A ARDs that bind histone H3 tails mono- or dimethylated at K9 (Extended Data Fig. 4a, b)¹⁰.

The structure predicts that methylation on H4K20 should break critical hydrogen bonds with the TONSL ARD. Isothermal titration calorimetry (ITC) and H4-tail peptide pull-downs confirmed that H4K20me1/2 is incompatible with TONSL binding (Fig. 2a–c). Furthermore, H4K20me2 significantly reduced binding of full-length recombinant TONSL–MMS22L to reconstituted mononucleosomes

(Fig. 2d). Recently, TONSL ARD with its neighbouring acidic stretch was proposed to bind H3K9me1 (ref. 5), but we were unable to detect an interaction between TONSL ARD (with or without the acidic stretch) and H3K9me1 peptides (Extended Data Fig. 4c, d). Together, our data show that TONSL binds to both free histones and nucleosomes via ARD recognition of H4 tails unmodified at K20 (Figs 1a–j, 2a–e and Extended Data Figs 1b, 2a–c, 3a–d). In line with this, H4K20me2 was not detected on TONSL-bound nucleosomal histones (Fig. 2f), while H4K16ac was present (Extended Data Fig. 5a). H4K16ac stimulated TONSL binding in peptide pull-downs (Fig. 2b, c and Extended Data Fig. 5b) and slightly enhanced ARD binding in ITC (Fig. 2a), but it did not overturn the inhibitory effect of H4K20me2 (Fig. 2e). However, H4K16ac is not essential for TONSL binding *in vivo*, as soluble histone H4 does not carry H4K16ac¹¹ and depletion of MOF, the major H4K16 acetyltransferase¹², did not significantly affect TONSL binding to chromatin (Extended Data Fig. 5c, d). In contrast, depletion of the H4K20 methyltransferase SET8 (also known as PR-SET7) significantly enhanced TONSL binding to chromatin in G1 cells in which H4K20me2 peaks^{13–16} (Fig. 2g and Extended Data Fig. 5e, f). We conclude that TONSL ARD is a histone-reader domain specific for H4 tails unmethylated at K20.

Given that TONSL–MMS22L binds new histones (devoid of H4K20me^{11,17}) in a pre-deposition complex with ASF1 and MCM2 (Fig. 1j and Extended Data Fig. 1a)¹, TONSL–MMS22L could be loaded onto replicating DNA together with new histones. To test how long after deposition new histones remain unmethylated at H4K20 with the potential to bind TONSL, we extracted H4K20 data from our recent large-scale proteomic study¹⁸, tracking modifications on new and old recycled histones by nascent chromatin capture (NCC)¹⁹ (Fig. 3a and Extended Data Fig. 6a, b). In nascent chromatin, new histones were exclusively unmethylated at H4K20 (98% H4K20me0), while old recycled histones were almost fully methylated at H4K20 (me1, 7%; me2, 88%; me3, 2%). Consistent with previous work^{13–16}, our analysis of primary cells (Extended Data Fig. 6c) and degradation of SET8 in S phase^{15,16}, new histones became methylated in late G2/M, rendering G1 chromatin devoid of H4K20me0 (Fig. 3a). This identifies H4K20me0 on new histones as a signature of post-replicative chromatin, implying that TONSL–MMS22L can bind H4 tails on new histones at replication forks and sister chromatids until late G2/M. Confirming this prediction, TONSL accumulated on chromatin in S phase, remained chromatin-bound in a population of G2 cells, and was excluded from chromatin in G1 (Fig. 3b and Extended Data Fig. 6d–f). To discriminate pre- and post-replicative chromatin, we labelled replicating DNA with 5-ethynyldeoxyuridine (EdU; pulse to mark ongoing replication, continuous labelling to identify post-replicative chromatin) and stained pre-replicative chromatin with MCM2 (refs 20, 21), and analysed colocalization with TONSL. TONSL staining was mutually exclusive with MCM2 (Fig. 3c and Extended Data Fig. 7a), but colocalized with EdU pulse labelling in very early S phase and with replicated DNA (continuous EdU labelling) throughout S phase (Fig. 3d and Extended Data Fig. 7b, c). TONSL was present at sites of ongoing DNA replication throughout S phase, but the degree of colocalization declined in mid/late S phase (Fig. 3d, left), consistent with TONSL binding to post-replicative chromatin also after fork passage (Fig. 3d, right). Mutation of the TONSL ARD abrogated recruitment of TONSL to chromatin, including DNA replication sites (Fig. 3e, f and Extended Data Fig. 7d–g). Together, these data demonstrate that TONSL is recruited to replication forks and post-replicative chromatin via ARD recognition of H4K20me0 on new histones.

Mutation of the TONSL ARD also abrogated chromatin binding and recruitment to replication forks in the presence of replication poisons such as camptothecin (CPT) and hydroxyurea (Fig. 4a–c). Furthermore, ARD mutation prevented accumulation of TONSL at site-specific double-strand breaks (DSBs; Fig. 4d and Extended Data Fig. 8a) and microlaser-generated DNA damage (Fig. 4e and Extended Data Fig. 8b, c). Co-staining with cell cycle markers confirmed that

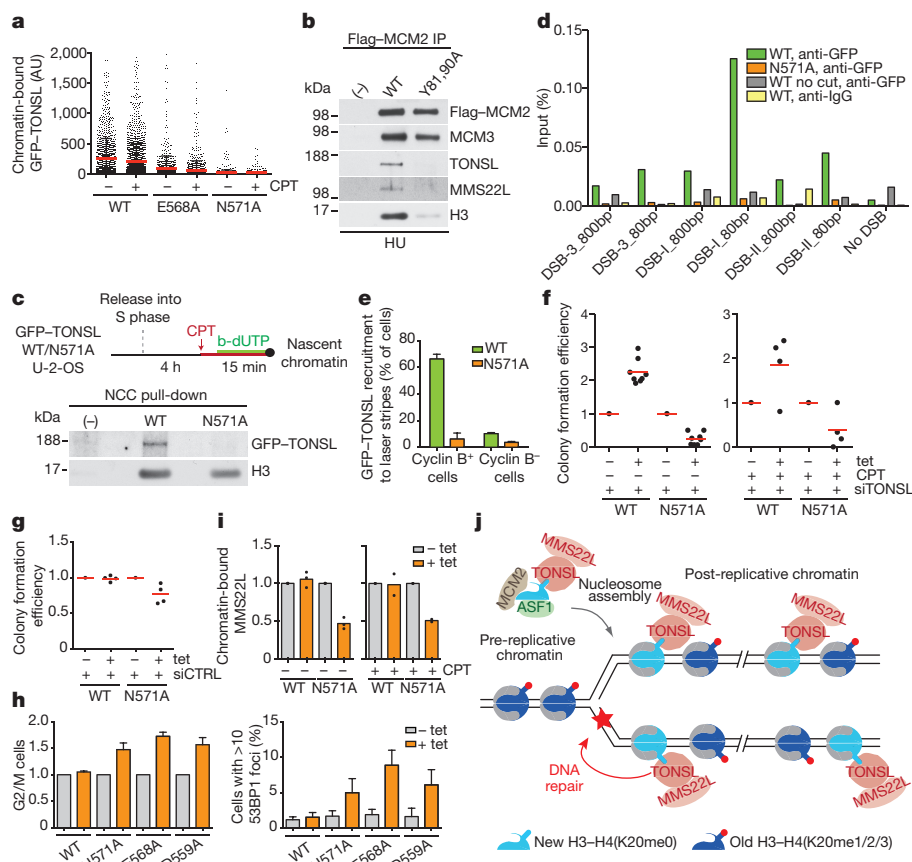


Figure 4 | H4K20me0 recognition is required for TONSL accumulation at DNA repair sites and genome stability. **a**, Chromatin-binding of GFP-TONSL in CPT-treated S-phase cells. WT, wild type. Error bars indicate s.d.; from left, $n = 1,461, 2,631, 1,245, 1,764, 2,116, 3,178$. **b**, Co-immunoprecipitation (IP) of TONSL-MMS22L with Flag-HA-MCM2 wild type or histone-binding mutant (Y81A, Y90A)⁷ from chromatin after hydroxyurea (HU) treatment. **c**, NCC analysis of GFP-TONSL recruitment to replication forks in CPT-treated cells. Minus sign indicates no b-dUTP. **d**, Chromatin immunoprecipitation (ChIP) and quantitative polymerase chain reaction (qPCR) analysis of GFP-TONSL recruitment to site-specific DSBs induced by AsiSI²⁹. See Extended Data Fig. 8a for additional controls. **e**, GFP-TONSL recruitment to laser-induced DNA lesions (error bars indicate s.d.; $n = 3$; total cells counted, 210 (wild type)

and 252 (N571)). **f**, **g**, Colony formation upon GFP-TONSL induction by tetracycline (+tet) in siRNA- and CPT-treated cells. **h**, Cell cycle and 53BP1 foci analysed by microscopy. Left, percentage of G2/M cells shown relative to non-induced cells (–tet). Error bars indicate s.d., $n = 4$ (left), 5 (right). **i**, Chromatin-bound MMS22L analysed as in Fig. 3e. Mean with individual data points are shown ($n = 3$ (untreated), 2 (CPT)), see Extended Data Fig. 8i for western blots. **j**, TONSL-MMS22L identifies post-replicative chromatin by binding H4K20me0 on new histones, directing TONSL-MMS22L genome surveillance function to DNA having a sister chromatid. Data are representative of three (**a**), two (**b–d**, **f**, right, **g**), and four (**f**, left) independent experiments. For protein inputs, see Extended Data Fig. 9e, f.

TONSL is recruited to DNA repair sites only in S and G2 cells, as expected² (Fig. 4e and Extended Data Fig. 8d, e). We conclude that H4K20me0 binding is required for TONSL accumulation at damaged forks and DNA lesions in post-replicative chromatin. However, this was not due to increased H4K20me0 (Extended Data Fig. 8f), suggesting that unmasking of H4 tails upon chromatin decompaction^{9,22} and/or interaction with repair factors contribute to TONSL-MMS22L accumulation at repair sites. Consistent with an auxiliary mode of recruitment, MMS22L interaction with RAD51 can stabilize the complex at challenged forks (P. Cejka and M. Peter, personal communication). Our data suggest that this is subsequent to H4K20me0 binding (Fig. 4a–e), and we thus next addressed the contribution of H4K20me0 recognition to TONSL-MMS22L function. In complementation analysis, TONSL wild type partially rescued the viability of TONSL-depleted cells in the presence and absence of CPT (Fig. 4f and Extended Data Fig. 8g, h), whereas TONSL ARD mutants were toxic (Fig. 4f and Extended Data Fig. 8g, h). In control cells, TONSL ARD mutants also reduced viability, causing G2/M arrest accompanied by replication-associated DNA damage (Fig. 4g, h). Furthermore, TONSL ARD mutants titrated MMS22L away from chromatin (Fig. 4i and Extended Data Fig. 8i), explaining the dominant-negative phenotype that mimics TONSL-MMS22L depletion^{1–4}. Collectively, this indicates that recognition of

H4K20me0 is central to TONSL-MMS22L function in safeguarding genome stability.

This study reveals that post-replicative chromatin has a distinct histone modification signature, read by the TONSL-MMS22L effector protein (Fig. 4j). This opens a new avenue to understand how DNA repair and other chromosomal transactions can be directly linked to the replication state of a genomic locus. Intriguingly, it is the new histones that make post-replicative chromatin distinct, and in this way H4K20me0 resembles the behaviour of H3K56ac²³ in yeast. Our data indicate that TONSL-MMS22L is delivered to nascent chromatin with new histones via the pre-deposition complex with MCM2 and ASF1 (Fig. 4j). We favour the idea that TONSL has a dual function as a histone chaperone⁵ and histone reader. Our structural work proposes that TONSL acts in a histone chaperone-like capacity by sequestering the H4 tail to prevent spurious contacts with DNA during H3–H4 deposition. Furthermore, TONSL ARD may counteract chromatin compaction by preventing association of the H4 tail with the H2A–H2B acidic patch on neighbouring nucleosomes. Thus, TONSL changes our perception of a histone chaperone by binding both soluble and nucleosomal histones. In its function as a histone reader, TONSL localizes MMS22L to post-replicative chromatin via H4K20me0 and allows TONSL-MMS22L to accumulate at damaged forks and DNA

lesions. We envision that H4K20me0 works as an affinity trap, making TONSL–MMS22L readily available to support RAD51 loading during homologous recombination. This provides a new approach and opportunity to understand the role of H4K20 in DNA repair, complementing the well-described role of H4K20me1/2 in recruiting 53BP1 to promote non-homologous end joining in competition with BRCA1–BARD1 (refs 24, 25). In post-replicative chromatin, H4K20me1/2 on old histones will support 53BP1 recruitment. Whether H4K20me0 on new histones also influences DNA repair pathway choice will be of interest in future investigations. It is notable that the structure of the TONSL ARD, including the histone-binding surface, is highly similar to the ARD of BARD1 (Extended Data Fig. 9a)²⁶, required for BRCA1 tumour suppressor function and homologous recombination²⁷. Multiple mutations in the TONSL ARD are reported in cancer (C608G, P557S, E597K; <http://cancer.sanger.ac.uk>) and the N571 residue, key to histone H4 binding, corresponds to the BARD1 N470S cancer mutation^{26,28}. This highlights the tumour suppressor function of H4K20me0 recognition, and the possibilities it brings for targeted cancer therapy should be explored in the future.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August 2015; accepted 5 May 2016.

Published online 22 June 2016.

- Duro, E. *et al.* Identification of the MMS22L–TONSL complex that promotes homologous recombination. *Mol. Cell* **40**, 632–644 (2010).
- O'Donnell, L. *et al.* The MMS22L–TONSL complex mediates recovery from replication stress and homologous recombination. *Mol. Cell* **40**, 619–631 (2010).
- O'Connell, B. C. *et al.* A genome-wide camptothecin sensitivity screen identifies a mammalian MMS22L–NFKBIL2 complex required for genomic stability. *Mol. Cell* **40**, 645–657 (2010).
- Piwko, W. *et al.* RNAi-based screening identifies the Mms22L–Nfkbil2 complex as a novel regulator of DNA replication in human cells. *EMBO J.* **29**, 4210–4222 (2010).
- Campos, E. I. *et al.* Analysis of the histone H3.1 interactome: a suitable chaperone for the right event. *Mol. Cell* **60**, 697–709 (2015).
- Groth, A. *et al.* Regulation of replication fork progression through histone supply and demand. *Science* **318**, 1928–1931 (2007).
- Huang, H. *et al.* A unique binding mode enables MCM2 to chaperone histones H3–H4 at replication forks. *Nature Struct. Mol. Biol.* **22**, 618–626 (2015).
- Richet, N. *et al.* Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork. *Nucleic Acids Res.* **43**, 1905–1917 (2015).
- Kalashnikova, A. A., Porter-Goff, M. E., Muthurajan, U. M., Luger, K. & Hansen, J. C. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J. R. Soc. Interface* **10**, 20121022 (2013).
- Collins, R. E. *et al.* The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nature Struct. Mol. Biol.* **15**, 245–250 (2008).
- Jasencakova, Z. *et al.* Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol. Cell* **37**, 736–743 (2010).
- Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25**, 6798–6810 (2005).
- Rice, J. C. *et al.* Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev.* **16**, 2225–2230 (2002).
- Pesavento, J. J., Yang, H. & Kelleher, N. L. & Mizzen, C. A. Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. *Mol. Cell. Biol.* **28**, 468–486 (2008).
- Beck, D. B., Oda, H., Shen, S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev.* **26**, 325–337 (2012).
- Jørgensen, S., Schotta, G. & Sørensen, C. S. Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res.* **41**, 2797–2806 (2013).
- Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol. Cell* **24**, 309–316 (2006).
- Alabert, C. *et al.* Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev.* **29**, 585–590 (2015).
- Alabert, C. *et al.* Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nature Cell Biol.* **16**, 281–293 (2014).
- Prasanth, S. G., Méndez, J., Prasanth, K. V. & Stillman, B. Dynamics of pre-replication complex proteins during the cell division cycle. *Phil. Trans. R. Soc. B* **359**, 7–16 (2004).
- Takahashi, T. S., Wigley, D. B. & Walter, J. C. Pumps, paradoxes and ploughshares: mechanism of the MCM2–7 DNA helicase. *Trends Biochem. Sci.* **30**, 437–444 (2005).
- Seeber, A., Hauer, M. & Gasser, S. M. Nucleosome remodelers in double-strand break repair. *Curr. Opin. Genet. Dev.* **23**, 174–184 (2013).
- Burgess, R. J. & Zhang, Z. Histone chaperones in nucleosome assembly and human disease. *Nature Struct. Mol. Biol.* **20**, 14–22 (2013).
- Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
- Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nature Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
- Fox, D., III *et al.* Crystal structure of the BARD1 ankyrin repeat domain and its functional consequences. *J. Biol. Chem.* **283**, 21179–21186 (2008).
- Laufer, M. *et al.* Structural requirements for the BARD1 tumor suppressor in chromosomal stability and homology-directed DNA repair. *J. Biol. Chem.* **282**, 34325–34333 (2007).
- Ishitobi, M. *et al.* Mutational analysis of BARD1 in familial breast cancer patients in Japan. *Cancer Lett.* **200**, 1–7 (2003).
- Iacovoni, J. S. *et al.* High-resolution profiling of γ H2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29**, 1446–1457 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the beam staff at the synchrotrons at the Argonne National Laboratory (NE-CAT) for technical assistance. We thank J. Rouse, D. Durocher, G. Legube and C. Storgaard Sørensen for reagents, G. Montoya for assistance with circular dichroism, C. B. Strømme, A. Strandsby, K. Nakamura, S.-b. Lee and M. Hödl for help with experiments, and Y. Antoku for assistance with microscopy. We thank J. Lukas for comments on the manuscript and Z. Jasencakova for illustrations. G.S. was supported by European Commission Marie Curie ITN FP7 'aDDRess'. D.J.P. was supported in part by grants from the Leukemia and Lymphoma Society and the STARR foundation. A.G. is an EMBO Young Investigator and her research is supported by the European Research Council (ERC StG, no. 281765), the Danish National Research Foundation to the Center for Epigenetics (DNRF82), the Danish Cancer Society, the Danish Medical Research Council, the Novo Nordisk Foundation and the Lundbeck Foundation. A.I. is supported by the European Commission FP7 Network of Excellence EpiGeneSys (project 257082), the DFG Excellence Clusters CIPSM and SyNergy, as well as the DFG Collaborative Research Center 1064 (projects A3 and Z3). T.B. is supported by the Medical Research Council and the European Research Council (ERC StG, no. 309952).

Author Contributions G.S. and A.G. conceived and led the functional studies. H.H. conceived and led the generation of cassette to crystallize the complex, and H.H. solved the structure and performed the ITC under the supervision of D.J.P. C.M.H. performed peptide pull-downs with ARD and recombinant nucleosomes. C.A. performed SET8 experiments and NCC. S.B.-J. and N.M. analysed recruitment to laser-induced DNA damage. N.R.-G. prepared histones for mass spectrometry and performed ChIP analysis. I.F. analysed histone modifications by mass spectrometry under the supervision of A.I. B.M.F. and T.B. prepared modified recombinant nucleosomes. L.M. and P.C. prepared recombinant TONSL–MMS22L. G.S., H.H., D.J.P. and A.G. wrote the manuscript and all authors commented on the manuscript.

Author Information Coordinate and structure factors have been deposited in the Protein Data Bank under accession number 5JA4. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J.P. (pateld@mskcc.org) or A.G. (anja.groth@bric.ku.dk).

Reviewer Information Nature thanks T. Kutateladze and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Protein expression and purification. No statistical methods were used to pre-determine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. All proteins used in this study, unless otherwise indicated, were expressed in BL21(DE3)-RIL cell strain (Stratagene). The human TONSL ARD (residues 512–692) and MCM2 HBD (fragments 61–130) were covalently linked through a four-glycine linker (G₄ linker) into one expression cassette. The MCM2 HBD–G₄–TONSL ARD expression cassette was cloned into a modified RSFDuet-1 vector (Novagen), with an N-terminal His₆-SUMO tag. The resulting plasmid was co-expressed with plasmid harbouring histone genes H3.3(Δ56) and H4. The expressed protein complex was first purified on an Ni-NTA affinity column. After removing the His₆-SUMO tag by using Ulp1 (SUMO protease), the protein complex was further purified on HiLoad 16/600 Superdex 200 column (GE Healthcare).

The GST-tagged TONSL ARD and its mutants including E530A, D559A, W563A, E568A, N571A and D604A were cloned into pGEX-6P-1 vector (GE Healthcare). The expressed proteins were first purified using Glutathione Sepharose 4B, then further purified by gel-filtration step. In some case, the GST tag was removed with 3C protease before the gel-filtration step. For purification of GST–H3 tail and GST–H4 tail proteins, the human histones H3 fragment 1–59 and H4 fragment 1–31 were cloned into pGEX-6P-1 vector respectively. The proteins were expressed and purified in the same way.

For production of recombinant full-length TONSL–MMS22L heterodimer, the sequence coding for full-length MMS22L was fused with an MBP tag at the 5' end and 10× His tag at the 3' end. The sequence coding for full-length TONSL was fused with a GST tag at the 5' end. Both MMS22L and TONSL constructs were cloned into a pFastBac1 vector. The complex was expressed in Sf9 cells by co-infection with both recombinant baculoviruses according to manufacturer's recommendation (Invitrogen). The proteins were extracted from Sf9 cells and purified similarly as described previously for Sgs1 (ref. 30). Briefly, the complex was purified on amylose resin, and MBP and GST tags were subsequently cleaved with PreScission protease. The heterodimer was then further purified using a Ni-NTA affinity resin. Washes were performed with 300 mM NaCl buffer.

Crystallization. At first, we tried to crystallize TONSL ARD in complex with a H4 tail or H3–H4 tetramer, but failed even with extensive screening. An additional binding protein may help to stabilize the whole complex and help crystallization. Then we tried to crystallize TONSL ARD in complex with the MCM2 HBD and H3–H4 tetramer. We just got very tiny crystals for this complex, but failed to get big and well-diffracted crystals. We realized that the whole complex of TONSL ARD with MCM2 HBD and H3–H4 tetramer might be destabilized by the harsh crystallization conditions and form a subcomplex, thus hindering the optimization of the crystals. Then we tried to covalently link TONSL ARD and MCM2 HBD into one cassette through different length of glycine linker (G_n linker). The G₁₂, G₁₁, G₁₀, G₉, G₈, G₇, G₆, G₅ and G₄ linkers had been tried and all these cassettes could be crystallized. One of the constructs with a G₄ linker gave well-diffracted crystals.

The G₄ linker complex, MCM2–HBD–G₄–TONSL–ARD cassette–H3.3(Δ56)–H4 complex (herein denoted as TONSL–ARD–MCM2–HBD–H3–H4 tetramer complex) at a concentration of 23 mg ml^{−1} was crystallized in 0.1 M MES pH 5.6, 7% isopropanol using sitting-drop vapour-diffusion method at 20°C. All the crystals were soaked in a cryoprotectant made from mother liquor supplemented with 25% glycerol before flash freezing in liquid nitrogen.

Structure determination. The data sets for the TONSL–ARD–MCM2–HBD–H3–H4 tetramer complex were collected at 0.979 Å on 24-ID-C/E NE-CAT (Advanced Photo Source, Argonne National Laboratory). All the data sets were processed by using the HKL 2000 program. The initial structure for the TONSL–ARD–MCM2–HBD–H3–H4 tetramer complex was solved by molecular replacement in PHASER³¹ with our previous structure of the MCM2–HBD–H3–H4 tetramer complex⁷ as a search model and manually refined and built using Coot³². The final structure of this complex was refined to 2.43 Å resolution using PHENIX³³. The Ramachandran plot showed 95.9% favoured and 4.1% allowed. Extended Data Table 1 summarizes the statistics for data collection and structural refinement.

Preparation of recombinant modified mononucleosomes. Recombinant human histone proteins were expressed in *Escherichia coli* BL21(DE3)-RIL cells from pET21b(+) (Novagen) vectors and purified by denaturing gel filtration and ion-exchange chromatography essentially as described³⁴. All histone proteins were dialysed into water containing 1 mM dithiothreitol (DTT), lyophilized and stored dry at −80°C. Modified H4 proteins were generated by native chemical ligation essentially as described for H3 (ref. 35). Briefly, tail-less H4 Δ1–28 I29C protein was expressed in *E. coli* BL21(DE3)-RIL cells from pET24b(+) (Novagen) and purified by denaturing gel filtration and reversed-phase chromatography using a ResourceRPC column (GE Healthcare). Purified H4 Δ1–28 I29C was then ligated to N-terminally acetylated H4 1–28 thioester peptides (Almac) and full-length ligated H4 was separated from unligated H4 Δ1–28 I29C by reversed-phase

chromatography via a C18 column (Aquapore RP-300/Perkin Elmer) using a gradient from 35% B to 45% B over 20 column volumes (A: 0.1% TFA in water; B: 90% acetonitrile; 0.1% TFA). Ligated H4 was directly lyophilized and stored dry at −80°C. Ligated H4 was refolded into octamers together with purified histones H2A, H2B and H3.1 and then assembled into nucleosomes with biotinylated 601-DNA as described^{34,35}.

GST pull-downs. For pull-downs of GST–ARD and its mutants including E530A, D559A, W563A, E568A, N571A and D604A with H3–H4, first 25 μl of Glutathione Sepharose 4B beads were suspended with 200 μl of binding buffer (20 mM Tris pH 7.5 and 0.5 M NaCl), and 1 nmol of GST–ARD proteins were added and incubated at 23°C for 10 min; then 0.5 nmol of pre-purified H3/H4 tetramers were added and incubated for another 1 h; then the beads were washed quickly with five times 1 ml of washing buffer (binding buffer, 1% Triton X-100) before adding 50 μl of sample loading buffer. An aliquot of 20 μl of each sample was analysed with SDS–PAGE. The GST pull-downs of histone tails of GST–H3_{1–59} and GST–H4_{1–31} with TONSL ARD were performed similarly.

Circular dichroism. Circular dichroism spectra were acquired using a Jasco J-815 Circular Dichroism Spectropolarimeter with a 1 mm quartz cuvette. Spectra were recorded for wild-type and mutant TONSL ARD (512–692, 6.25 μM) between 260 nm and 195 nm in KH₂PO₄/K₂HPO₄ buffer (25 mM, pH 7.8) with a data pitch of 0.5 nm, bandwidth of 1 nm and with three accumulations at a scanning speed of 50 nm min^{−1}.

In vitro translation and pull-downs with H3–H4 sepharose beads. NHS-activated sepharose 4 Fast-Flow beads (GE Healthcare) were washed with 0.1 M HCl and incubated overnight with 1 μM recombinant histone H3.1–H4 tetramers (New England Biolabs, catalogue number M2509S) or 1 μM recombinant histone H2A–H2B dimers (New England Biolabs, catalogue number M2508S) in Coupling Buffer (0.2 M NaHCO₃, 0.2 M NaCl). One microgram of pSC-B-TONSL, pEXPR-IBA-105-ASF1A wild-type and pEXPR-IBA-105-ASF1A V94R plasmids was incubated with TnT Quick Coupled Transcription/Translation System (Promega) and ³⁵S-methionine according to the manufacturer's instructions. Ten microlitres of *in vitro* translation (IVT) mixture were added to the H3.1–H4, or H2A–H2B, sepharose beads and incubated for 2 h. Beads were washed with 200 mM NaCl, 0.2% NP40 buffer. Beads were boiled in 1× LSB and loaded on a 4–12% Bis-Tris NuPage gel (Life Technologies). Proteins were transferred to a 0.2 μm nitrocellulose membrane by overnight wet transfer at 20 V and the membrane was incubated in an autoradiography cassette for 24 h before detection by Phosphor Imager (PerkinElmer).

ITC experiments. All the ITC titrations were performed on a Microcal ITC 200 calorimeter at 25°C or 20°C. The peptides of H4 (residues 9–25) and its modified peptides K16ac (with acetylation on Lys16), H18W (with His18 mutated to Trp18), H4K20me1 (monomethylation on Lys20) and H4K20me2 (dimethylation on Lys20), and peptide of H3(1–21)K9me1 (monomethylation on Lys9) were all synthesized at Tufts University Core Facility. The exothermic heat of the reaction was measured by 17 sequential 2.2 μl injections of the peptides (1.41 mM in buffer 20 mM Tris pH 7.5 and 0.5 M NaCl) into 200 μl of the TONSL ARD solution (145 μM in the same buffer), spaced at intervals of 150 s or 180 s. The data were processed with Microcal Origin software and the curves were fit to a single site binding model.

Peptide pull-downs assays. Purified recombinant TONSL ARD (residues 512–692) was stored at 400 μM in 1 M NaCl, 20 mM Tris HCl pH 7.5 at −80°C. For each pull-down, 400 pmol of the ARD stock (1 μl, 400 μM) was diluted with 99 μl of binding buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 5% glycerol, 0.25% NP-40, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 1 mM leupeptin, 1 mM pepstatin). ARD input material was scaled to the number of pull-downs performed. For each pull-down, an H4 peptide (JPT Peptide Technologies GmbH) spanning residues 14–33 (2.5 μl, 250 μM) with a C-terminal biotinoyl-lysine residue or, as control, biotin (2.5 μl, 400 μM) was added to 1.1 ml of binding buffer in addition to 100 μl of the ARD input material and the mixture was incubated overnight rotating at 4°C. The next day, 25 μl of MyOne Streptavidin C1 beads (Life Technologies) was washed in binding buffer (3 × 500 μl) for each pull-down, removing the final wash from the beads. The ARD/peptide or ARD/biotin mixture was added to an aliquot of pre-washed MyOne Streptavidin C1 beads and incubated with rotation at 4°C for 3 h. Finally, the beads were washed (2 × 300 μl and 1 × 200 μl of 300 mM NaCl, 50 mM Tris HCl pH 7.5, 5% glycerol, 0.25% NP-40, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 1 mM leupeptin, 1 mM pepstatin) and pull-down material was visualized by Coomassie staining after SDS–PAGE separation of proteins on a NuPAGE 4–12% gel.

For pull-downs from cell extracts, MyOne T1 beads were incubated overnight with 1 μg of biotinylated peptides in high salt (HS; 300 mM NaCl, 0.5% NP40, Tris HCl, EDTA, 5% glycerol) buffer and subsequently washed twice with PBS. One milligram of NP40/NaCl extract from HeLa S3 or GFP–TONSL U-2-OS cells was added to the beads and incubated for 2 h rotating at 4°C. The beads were

then washed five times with HS buffer, 2 min rotating at 4°C. After washing, the beads were resuspended in 1× LSB and boiled for 10 min. The eluted proteins were loaded on a 4–12% Bis-Tris NuPage gel (Life Technologies). Proteins were then transferred to a 0.2 µm nitrocellulose membrane by overnight wet transfer at 20 V and detected by western blotting.

Nucleosome pull-down assay. Modified nucleosomes of H4K20me0 or H4K20me2 were prepared by peptide ligation and stored at 0.1 µg µl⁻¹ (by histone octamer) in 100 mM NaCl, 50 mM Tris HCl pH 7.5 at 4°C. Full-length TONSL–MMS22L complex was stored at 746 nM in 100 mM NaCl, 50 mM Tris HCl, pH 7.5, 5 mM β-mercaptoethanol, 10% glycerol, 0.5 mM PMSF at –80°C. Nucleosome pull-downs were performed across two sets of conditions ($n = 3$ for each condition) in the presence of herring sperm competitor DNA. Condition number 1: Nucleosomes (1 µg by histone octamer) or biotin (0.5 µg) were mixed with TONSL–MMS22L (1.9 pmol) and made up to 30 µl with binding buffer (500 mM NaCl, 50 mM Tris HCl pH 7.5, 20% glycerol, 0.1% NP-40, 1 mM DTT, protease inhibitors and 10 µg ml⁻¹ herring sperm DNA (Sigma)). Inputs of 10 µl were taken before diluting each sample with binding buffer to a final volume of 300 µl and incubating overnight at 4°C. Pull-downs were performed by adding 20 µl of MyOne Streptavidin C1 beads prewashed and resuspended in 100 µl of binding buffer to each pull-down reaction, incubating at 4°C for 2 h, washing with 5 × 500 µl binding buffer for 2 min at room temperature. Condition number 2: nucleosomes (0.5 µg by histone octamer) or biotin (0.5 µg) were mixed with TONSL–MMS22L (1.3 pmol) and made up to 30 µl with binding buffer (500 mM NaCl, 50 mM Tris HCl pH 7.5, 5% glycerol, 0.5% NP-40, 0.2 mM EDTA, 1 mM DTT, protease inhibitors and 10 µg ml⁻¹ herring sperm DNA (Sigma)). Inputs of 15 µl were taken before diluting each sample with binding buffer to a final volume of 500 µl and incubating overnight at 4°C. Pull-downs were performed by adding 10 µl of MyOne Streptavidin T1 beads prewashed and resuspended in 100 µl of binding buffer to each pull-down reaction, incubating at 4°C for 4 h, washing with 5 × 500 µl binding buffer for 2 min at room temperature. Pull-downs were visualized by SYPRO Ruby staining after SDS–PAGE separation of proteins on a NuPAGE 4–12% gel using an ImageQuant LAS 4000 (GE Healthcare). The intensity of stained bands were quantified using ImageJ, TONSL intensity was normalized to the combined intensity of H3, H2A and H2B. Statistical analysis was performed using data from the six independent experiments using the unpaired *t*-test with equal standard deviations in prism 6.

Cell culture, transfection and drug treatment. U-2-OS (gift from J. Bartek), HeLa S3 (gift from P. Nakatani) and TIG-3 (gift from K. Hansen) cells were grown in DMEM (Gibco) containing 10% FBS (Hyclone) and 1% penicillin/streptomycin and drugs for selection. The construct for siRNA-resistant GFP–TONSL was described² and ARD mutations were introduced in this construct by site-directed mutagenesis. The construct for pBABE–SNAP–HA–H4 plasmid was previously described³⁶ and H4 tail mutations were introduced in this construct by site-directed mutagenesis. Cells inducible for GFP–TONSL wild type and ARD mutants were generated in Flp-In T-Rex U-2-OS cells (Invitrogen) by transfection of pcDNA5/FRT/TO–GFP–TONSL plasmids with Lipofectamine 2000, according to the manufacturer's protocol, and selection with hygromycin (200 µg ml⁻¹). Previously described inducible GFP–TONSL U-2-OS cells¹ were used for Fig. 1h and Fig. 2c, e. U-2-OS Flag–HA–MCM2 wild type and Y81A and Y90A cells were previously described⁷. pBABE–AsiSI–ER–HA²⁹ was introduced into inducible GFP–TONSL cell lines by lentiviral infection and puromycin selection. All cell lines were authenticated by western blotting and/or immunofluorescence. All cell lines used in this study tested negative for mycoplasma contamination. Expression of GFP–TONSL was induced by addition of 1 µg ml⁻¹ of tetracycline for 24 h. U-2-OS and TIG3 cells were synchronized by a single thymidine block (2 mM) and released into S phase in the presence of 24 µM dCTP. For transient expression of GFP–TONSL or SNAP–HA–H4 (Fig. 1h, j), expression plasmids were introduced by transfection with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol and cells harvested 24 h after transfection. siRNA transfection was performed with RNAiMax reagent (Invitrogen) according to the manufacturer's protocol. All siRNAs were used to a final concentration of 50 nM. siRNA sequences (Sigma): siSET8#1: 5'-GUACGGAGCGCAUGAAGU-3'; siSET8#2: 5'-ACUUAUGGCGCUCGUAACUU-3' (ref. 37); siMOF#1: 5'-GUGAUCAGUCGAGUGA-3' (ref. 12); siMOF#2: 5'-GAGAUCAACCAUGUGCAGA-3'; siTONSL: 5'-GAGCUGGACUUAAGCAUGA-3' (ref. 2).

Drug treatment was as follows. CPT: cells were either treated with 1 µM CPT for 3 h (Fig. 4a, i and Extended Data Fig. 8f) or 20 min (Fig. 4c), or with 50 nM CPT for 24 h (Fig. 4f). Hydroxyurea: cells were treated with 3 mM hydroxyurea for 2 h (Fig. 4b) or 3 h (Extended Data Fig. 8f).

Cell extracts and chromatin solubilisation. For detergent/salt soluble extracts, HeLa S3 and U-2-OS cells were washed with cold PBS, scraped and incubated for 15 min on ice in HS buffer supplemented with trichostatin A (TSA) and protease and phosphatase inhibitors (5 mM sodium fluoride, 10 mM β-glycerolphosphate, 0.2 mM sodium vanadate, 10 µg ml⁻¹ leupeptin, 10 µg ml⁻¹ pepstatin, 0.1 mM

PMSF, Sigma). After centrifugation at 16,000g for 15 min at 4°C, the supernatant was collected. To analyse chromatin-bound complexes, cells were washed twice in cold PBS, scraped and centrifuged at 1,500g for 10 min at 4°C. The pellet was incubated on ice for 10 min in CSK buffer (10 mM PIPES pH 7, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂/0.5% Triton X-100, supplemented with TSA and protease and phosphatase inhibitors (5 mM sodium fluoride, 10 mM β-glycerolphosphate, 0.2 mM sodium vanadate, 10 µg ml⁻¹ leupeptin, 10 µg ml⁻¹ pepstatin, 0.1 mM PMSF, Sigma) and subsequently centrifuged at 1,500g for 10 min to collect soluble proteins. For DNaseI or benzonase release of chromatin material, the remaining pellet was resuspended in CSK/0.1% Triton X-100 containing DNase I (1,000 U ml⁻¹, Roche), or benzonase (2,500 U ml⁻¹, Millipore), and incubated at 30°C for 30 min. Solubilized chromatin was then collected by centrifugation at 16,000g for 10 min.

Immunoprecipitation from cell extracts. Immunoprecipitation was performed with agarose magnetic GFP–Trap beads (Chromotek), anti-Flag magnetic beads (Sigma) and anti-HA magnetic beads (Life Technologies). Cell extracts were incubated with beads for 2 h at 4°C rotating. The beads were subsequently washed five times with HS buffer and resuspended in 1× LSB before boiling and SDS–PAGE separation of proteins on a NuPAGE 4–12% gel.

Western blotting and antibodies. The following antibodies were used: TONSL (Abcam ab101898), TONSL (Sigma, HPA024679; validated in Extended Data Fig. 6d), MMS22L¹, H3 (Abcam ab1791, Abcam ab10799), GFP (Santa Cruz sc-8334, Abcam ab290), biotin (Abcam ab53494), MCM2 (BD Biosciences 610701), H2B (Abcam ab1790), H4K16ac (Millipore 07-329), H4K20me1 (Abcam ab9051), H4K20me2 (Cell Signaling 9759), 53BP1 (Santa Cruz sc-22760; Novus Biologicals NB100-904), γ-H2AX (Millipore 05-636), Cyclin B (BD Biosciences 610220), RPA70 (Abcam ab79398), SET8 (Millipore, 06-1304), MCM3 (Abcam ab 4460). Secondary antibodies conjugated with horseradish peroxidase (HRP) were from Jackson ImmunoResearch Labs. Signals were revealed by chemiluminescence substrate from Pierce (SuperSignal West Pico or SuperSignal West Femto).

FACS and analysis. For analysis of cell cycle progression, cells were fixed in 70% ethanol and stained with propidium iodide/RNase for 30 min in the dark, before analysis on a FACS Calibur machine. FACS profiles were analysed by FlowJo 10.0.8 software.

Mass spectrometry. Histones from TIG3 fibroblasts were extracted from chromatin as previously described¹⁸. Protein was resuspended in 50 µl of 100 mM triethylammonium bicarbonate (TEAB; Sigma), pH adjusted with 2 µl of 1.5 M Tris pH 8 and digested for 16 h at 37°C with 3 µl of 20 ng µl⁻¹ Asp-N (Wako) in 100 mM TEAB. After 15 min centrifugation at 10,000g at 25°C, the supernatant was placed in a new tube and digestion was repeated for the pellet during 4 h under the conditions described earlier. The digested peptides of both digestions were merged, acidified with 10 µl of 1% TFA and purified using sequential StageTip C18 and Carbon Toppit (Glygen). Purified peptides were evaporated, resuspended in 15 µl of 0.1% TFA. Injected material was normalized to analyse by liquid chromatography mass spectrometry (LC–MS) the histones corresponding to 9.0 × 10⁵ cells. The LC method was used as described elsewhere¹⁸. The MS was performed in an Orbitrap Classic with similar settings as described previously¹⁸ but with survey scan range at 550–690 *m/z* and MS2 set in scheduled and targeted data independent mode for the four-time charged ions of the four different methylation states (unmodified, mono-, di- and trimethylated H4K20). Peptides were quantified using the peak area from the corresponding extracted ion chromatograms (±10 p.p.m.).

Immunofluorescence, microscopy and laser microirradiation. U-2-OS cells conditional for GFP–TONSL were grown on glass coverslips or 96-well plates and either directly fixed in 4% paraformaldehyde (PFA) for 10 min or washed in CSK, pre-extracted 5 min with cold CSK/0.5% Triton X-100 and rinsed with CSK and PBS before fixation in 4% PFA for 10 min. Coverslips were mounted on glass slides with Mowiol mounting medium (Sigma-Aldrich) containing DAPI. Fluorescence images were collected on a DeltaVision system with a ×40 or ×60 oil immersion objective. For colocalization analysis by deconvolution microscopy, z-stacks were acquired (step of 0.2 µm), deconvolved and analysed by SoftWoRx 5.0.0. Pearson coefficient correlation analysis was performed on single cells using SoftWoRx 5.0.0. Brightness and contrast were adjusted using Adobe Photoshop CS6. For high-content quantitative analysis, fluorescence images were acquired using an Olympus ScanR high-content microscope and processed on the ScanR analysis software. More than 5,000 cells per sample were analysed. Cell cycle phases were gated on DAPI and EdU intensity. Graphs were generated with TIBCO Spotfire software. For microirradiation experiments, cells grown on glass coverslips were fixed in 4% formaldehyde for 15 min, permeabilized with PBS containing 0.2% Triton X-100 for 5 min and incubated with primary antibodies diluted in DMEM for 1 h at room temperature. After staining with secondary antibodies (Alexa Fluor 488, 568 and 647; Life Technologies) for 30 min, coverslips were mounted on glass slides in Vectashield mounting medium (Vector Laboratories) containing

the nuclear stain DAPI. For detection of nucleotide incorporation during DNA replication, an EdU-Plus labelling kit (Life Technologies) was used according to the manufacturer's instructions. Confocal images were acquired on an LSM-780 (Carl Zeiss) mounted on a Zeiss-AxioObserver Z1 equipped with a Plan-Neofluar $\times 40/1.3$ oil immersion objective. Image acquisition and analysis was carried out with LSM-ZEN software. Laser microirradiation of cells was performed essentially as described³⁸.

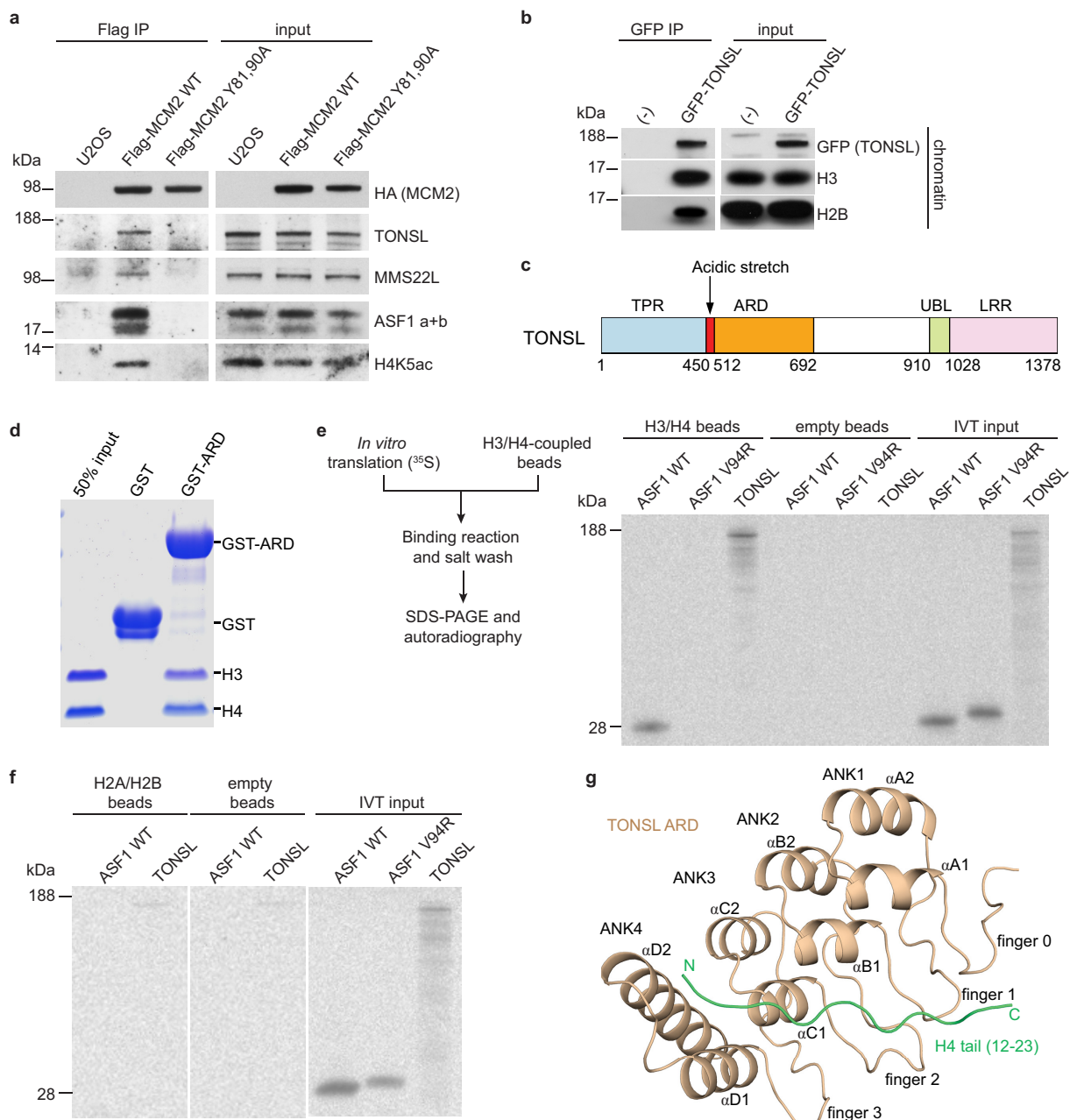
ChIP. GFP-TONSL wild type and N571A U-2-OS harbouring the inducible ER-HA-AsiSI endonuclease²⁹ were treated with 4-OHT and 10 μ M DNA-PK inhibitor NU7026 (Millipore) for 4 h to increase homologous recombination³⁹. Cells were cross-linked for 10 min in 1% formaldehyde and chromatin was fragmented by sonication using Bioruptor Sonicator (Diagenode). ChIP was performed as previously described⁴⁰ with the following modifications: 30 μ g of chromatin was immunoprecipitated with 5 μ g of anti-GFP (Abcam ab290) and rabbit-IgG. Immunoprecipitated DNA was analysed in duplicate by RT-qPCR. In all cases, γ H2A.X induction was verified by immunofluorescence and a sample without 4-OHT was included as a 'no cut' control. Primer pairs for the analysis of DSB-3, DSB-I and DSB-II are described⁴¹. Primer sequences used for the amplification of a genomic region devoid of DSBs were as follows: noDSB-for: 5'-TGACAAGGACAGGTCTTCC; noDSB-rev: 5'-CACCGTCCG TTGTATGTCTG. ChIP efficiency was calculated as percentage of input DNA immunoprecipitated.

NCC. The NCC protocol¹⁹ was adjusted for adherent U-2-OS cells. CPT (1 μ M) was added 5 min before b-dUTP labelling and was included in all steps until fixation. Cells were incubated for 5 min in a hypotonic buffer (50 mM KCl, 10 mM HEPES) containing b-dUTP and resuspended into fresh cell culture medium for an additional 15 min. Cells were fixed 15 min in 1% formaldehyde, rinsed twice in PBS and collected by scraping in cold room. Nuclei were mechanically isolated in sucrose buffer (0.3 M sucrose, 10 mM HEPES-NaOH at pH 7.9, 1% Triton X-100 and 2 mM MgOAc). Chromatin was solubilized by 28 cycles 30 s on, 90 s off in sonication buffer (10 mM HEPES-NaOH at pH 7.9, 100 mM NaCl, 2 mM EDTA at pH 8, 1 mM EGTA at pH 8, 0.2% SDS, 0.1% sodium sarkosyl and 1 mM phenylmethylsulfonylfluoride) using a Bioruptor at 4 °C. Solubilized chromatin was pre-cleared using streptavidin-coated magnetic beads (MyC1 Streptavidin beads) pre-incubated with biotin. b-dUTP labelled chromatin was next purified over night at 4 °C using streptavidin-coated magnetic beads. Beads were washed five times for 2 min in wash buffer (10 mM HEPES-NaOH pH 7.9; 200 mM NaCl; 2 mM EDTA pH 8; 1 mM EGTA pH 8; 0.1% SDS; 1 mM PMSF). Total chromatin (input) and isolated nascent chromatin were boiled for 40 min on beads in LSB 1 \times (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS, 8% glycerol, bromophenol blue)

and separated by SDS-PAGE for western blotting. Pulse-SILAC-NCC (Fig. 3a) was performed as described¹⁸.

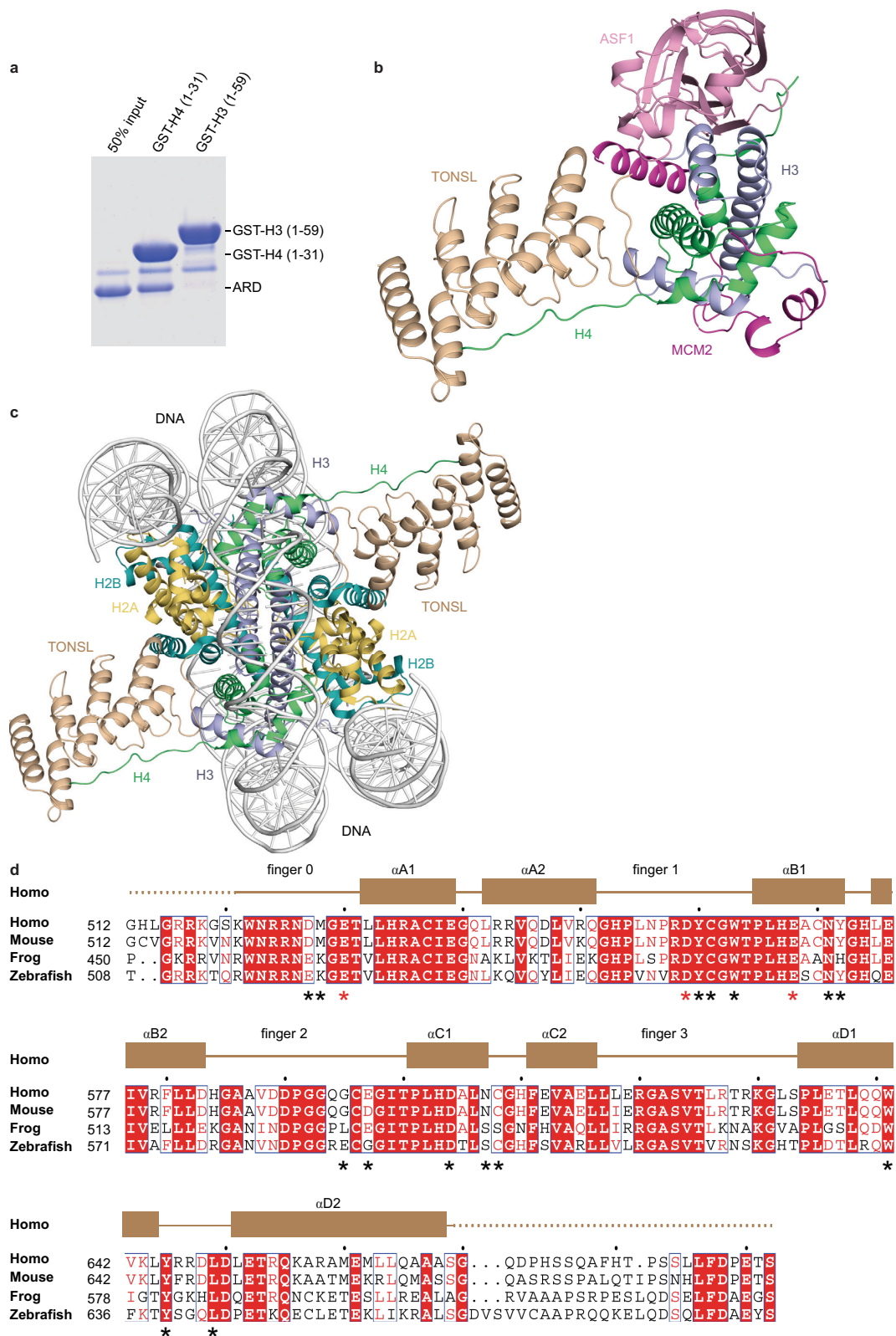
Clonogenic assay. U-2-OS inducible for GFP-TONSL ARD wild type and mutant were transfected with siRNA, trypsinized 24 h later and seeded in technical triplicates of 1,000 or 3,000 cells in the presence or absence of tetracycline. After 24 h, cells were washed to remove tetracycline and CPT was added for 24 h as indicated. Cells were then cultured in fresh medium for 12–15 days before fixation and staining with MeOH/Crystal Violet. Colony formation efficiency was determined by manual colony counting or quantification of Crystal Violet staining by ImageJ software and normalized to non-induced control. Each data point represents a technical triplicate of 1,000 or 3,000 seeded cells within each biological replicate.

30. Cejka, P. & Kowalczykowski, S. C. The full-length *Saccharomyces cerevisiae* Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. *J. Biol. Chem.* **285**, 8290–8301 (2010).
31. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
33. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
34. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
35. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484 (2010).
36. Bodor, D. L., Valente, L. P., Mata, J. F., Black, B. E. & Jansen, L. E. Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. *Mol. Biol. Cell* **24**, 923–932 (2013).
37. Jørgensen, S. *et al.* The histone methyltransferase SET8 is required for S-phase progression. *J. Cell Biol.* **179**, 1337–1345 (2007).
38. Mosbech, A., Lukas, C., Bekker-Jensen, S. & Mailand, N. The deubiquitylating enzyme USP44 counteracts the DNA double-strand break response mediated by the RNF8 and RNF168 ubiquitin ligases. *J. Biol. Chem.* **288**, 16579–16587 (2013).
39. Nakamura, K. *et al.* Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Mol. Cell* **41**, 515–528 (2011).
40. Jakobsen, J. S. *et al.* Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries. *Genome Res.* **23**, 592–603 (2013).
41. Aymard, F. *et al.* Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature Struct. Mol. Biol.* **21**, 366–374 (2014).



Extended Data Figure 1 | TONSL binding to histones *in vivo* and *in vitro*. **a**, Histones bridge the interaction between TONSL–MMS22L and MCM2 in cell extracts as shown by co-immunoprecipitation of Flag–HA–MCM2 wild type or histone-binding mutant (Y81A, Y90A)⁷. U-2-OS cell inducible for Flag–HA–MCM2 wild type or Y81A, Y90A⁷ were induced for 24 h before immunoprecipitation with Flag antibodies (one representative experiment out of two is shown). **b**, Immunoprecipitation of GFP–TONSL from solubilized chromatin of HeLa cells transiently transfected with GFP–TONSL plasmid, showing that TONSL associates with nucleosomal histones H3 and H2B (one representative experiment out of two is shown).

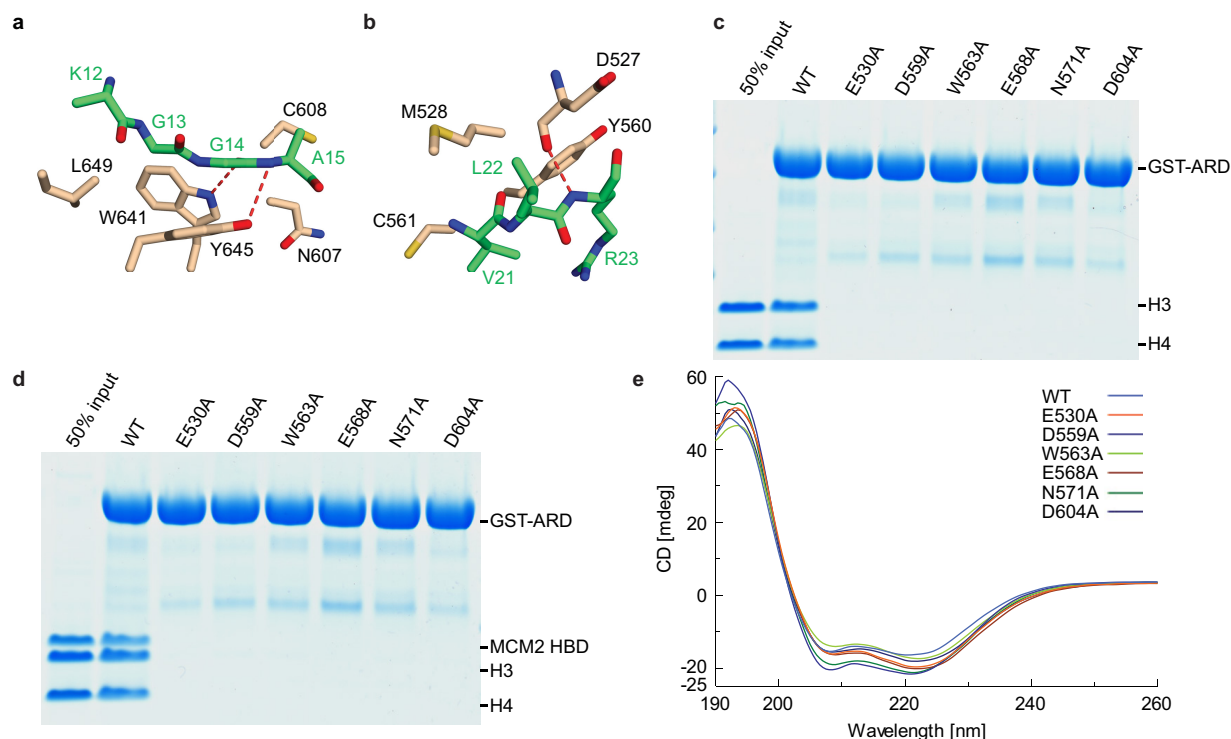
c, Domain structure of TONSL^{1–4}. LRR, leucine-rich repeats; TPR, tetratricopeptide repeats; UBL, ubiquitin-like domain. **d**, Pull-down of GST–ARD with recombinant histones H3–H4 tetramers. **e**, **f**, Pull-down of *in vitro*-translated full-length TONSL with recombinant histones H3–H4 tetramers (**e**) or H2A–H2B dimers (**f**) coupled to NHS-activated sepharose beads (one representative experiment out of three (**e**) and two (**f**) is shown). ASF1a wild type and histone-binding mutant (V94R) were included as controls. **g**, TONSL ARD consists of four ankyrin repeats and uses its elongated concave surface to target the H4 tail spanning residues 12 to 23.



Extended Data Figure 2 | See next page for caption.

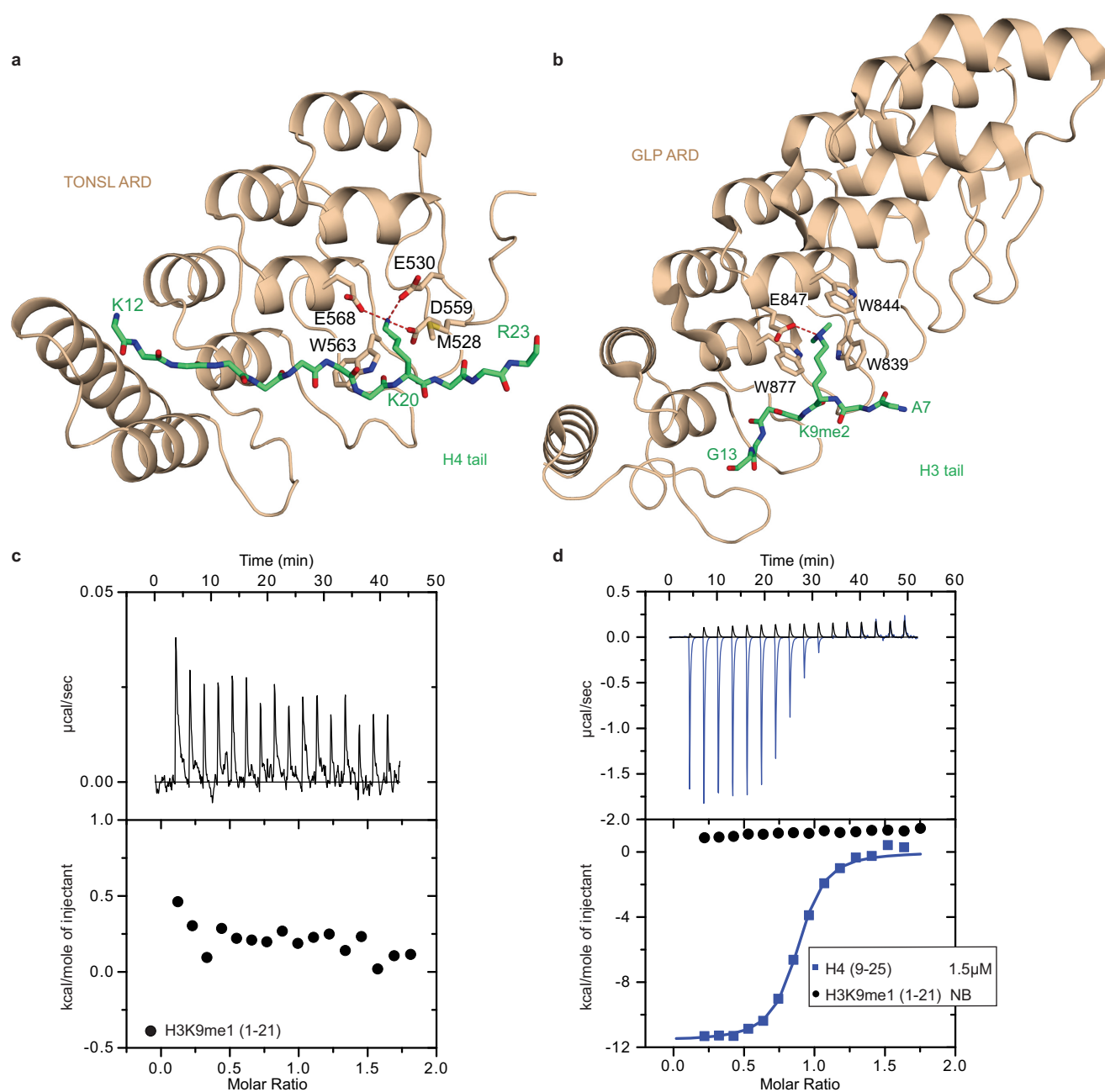
Extended Data Figure 2 | Models and sequence alignment of TONSL ARD. **a**, Pull-down assay of recombinant ARD with GST–H3 tail (amino acids 1–59) and GST–H4 tail (amino acids 1–31). **b**, Modelling of TONSL ARD on the co-chaperone structure of MCM2 HBD and ASF1 in complex with an H3–H4 dimer. When comparing the structure of the TONSL-ARD–MCM2-HBD–H3–H4 tetramer complex with our previous structure of the MCM2-HBD–H3–H4-dimer–ASF1 complex⁷ (Protein Data Bank accession 5BNX), the common parts of both structures superimposed well with a small root mean squared deviation (r.m.s.d.) of 0.44 Å. A model of the quinary complex composed of one molecule of each protein, TONSL ARD, MCM2 HBD, ASF1, H3 and H4, was made after superposition. This model shows that TONSL ARD, MCM2 HBD and ASF1 could simultaneously bind an H3–H4 dimer without steric

clash. **c**, Model of TONSL ARD on the structure of the nucleosome. The model was generated by a direct superposition of the H3–H4 tetramer in the structure of the TONSL ARD–MCM2 HBD–H3–H4 tetramer complex onto the H3–H4 tetramer in the nucleosome structure (Protein Data Bank accession 3AV2). There was no adjustment in the conformation of the model and no steric clash in the model. The MCM2 HBD molecules were omitted from the model for clarity. **d**, Alignment of TONSL ARD (512–692) sequences from *Homo sapiens*, *Mus musculus*, *Xenopus laevis* and *Danio rerio*. The secondary structures of human TONSL ARD are showed on top of the sequence alignment. Asterisks indicate the highly conserved residues that constitute the H4 tail-binding surface of TONSL ARD and the three strictly conserved acidic residues forming hydrogen bonds with the key residue H4 Lys20 are highlighted with red asterisks.



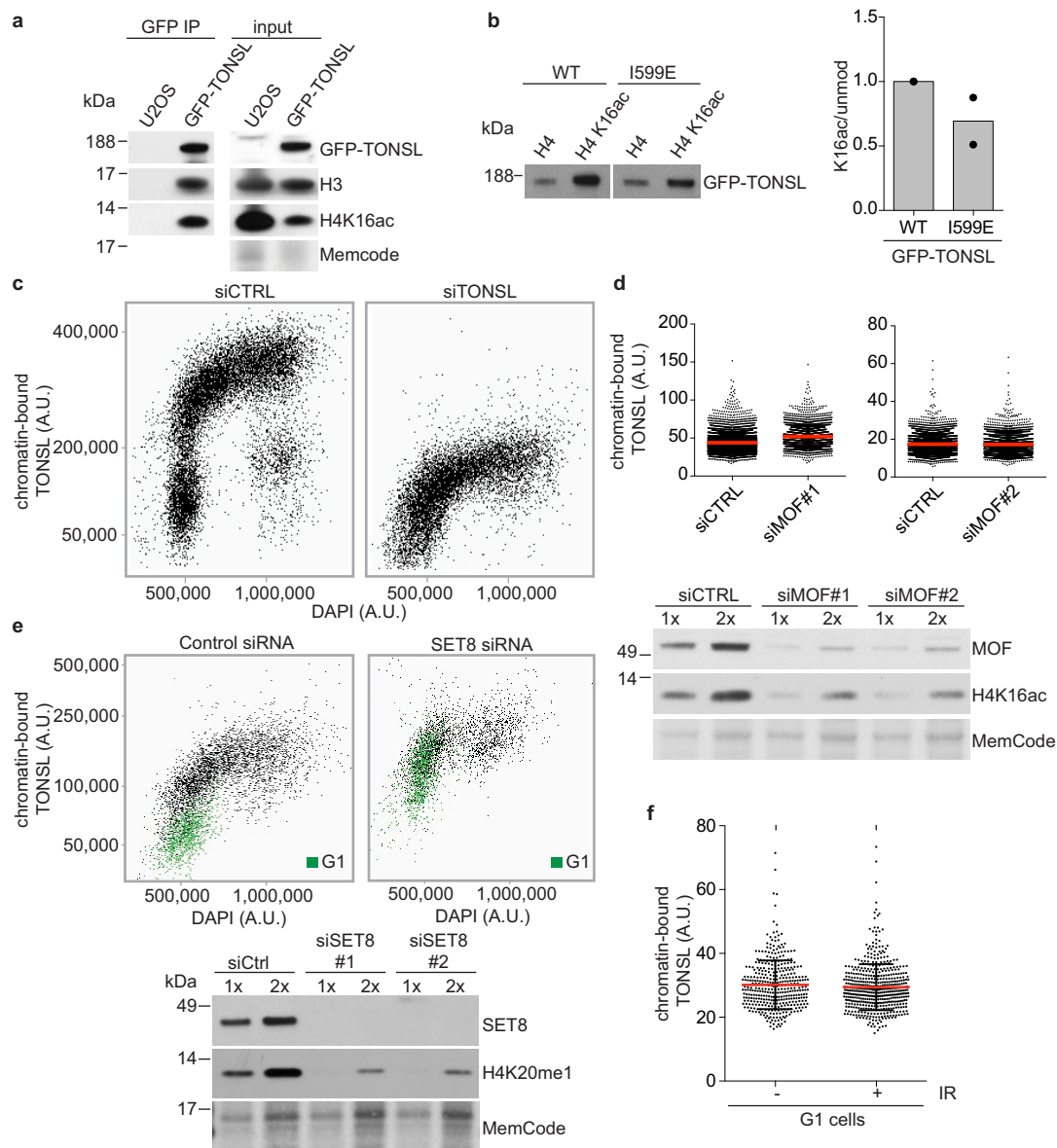
Extended Data Figure 3 | Interaction details of TONSL ARD and GST pull-downs. **a, b**, Molecular details of the interactions of TONSL ARD with H4 tail region residues 12–15 (**a**) and residues 21–23 (**b**). The Lys12–Gly13–Gly14–Ala15 segment of H4 is positioned within a narrow surface channel of the TONSL ARD scaffold. The intermolecular contacts spanning the Lys12–Gly13–Gly14–Ala15 segment of H4 include hydrophobic interactions between residues Gly13, Gly14 and Ala15 of H4 and residues Asn507, Cys508, Trp641, Tyr645 and Leu649 of ARD, as well as hydrogen bonds between the main-chain O of H4 Gly14 and N ϵ 1 of ARD Trp641, and between the main-chain N of H4 Ala15 and O η of ARD Tyr645 (**a**; Fig. 1c). The main-chain O of H4 Lys16 hydrogen bonds with the N δ 2 of ARD Asn571, while the side chain of H4 Lys16 forms contacts with ARD Asn607 and electrostatic interactions with the side chain of ARD Glu597 (Fig. 1c). The side chain of H4 Arg17 stacks over the side chains of ARD Tyr572 and Cys608, while its N η 1 atom forms two hydrogen bonds with main-chain O and O δ 1 of ARD Asn571 (Fig. 1c, e). The side chain of H4 H18 penetrates into a pocket lined by

four strictly conserved residues (Trp563, Glu568, Asn571 and Asp604) and is positioned over His567 of ARD (Fig. 1c, f). The side chain of H4 His18 is stacked between Trp563 and Asn571 and forms hydrogen bonds to Glu568 and Asp604 of ARD (Fig. 1f). The main-chain O of H4 Arg19 forms a hydrogen bond with N ϵ 1 of Trp563 and its side chain forms contacts with Cys561 and Gly595 of ARD (Fig. 1c). Interactions with the key residue H4 Lys20 are described in the text (Fig. 1g). The intermolecular contacts spanning the Val21–Leu22–Arg23 segment of H4 include contacts between side chains of H4 Val21 with Tyr560 and Cys561 of ARD (**b**), while H4 Leu22 interacts with Asp527 and Met528 of ARD. The main-chain N of H4 Arg23 forms a hydrogen bond with the main-chain O of Asp527 of ARD, while the side chain packs against the side chain of Tyr560 of ARD (**b**). **c**, Pull-down of recombinant histones H3–H4 with GST–TONSL ARD wild type or indicated mutants. **d**, Pull-down of pre-purified MCM2 HBD–H3–H4 tetramer complex with GST–TONSL ARD wild type or indicated mutants. **e**, Circular dichroism analysis of TONSL ARD wild type and the indicated ARD mutants.



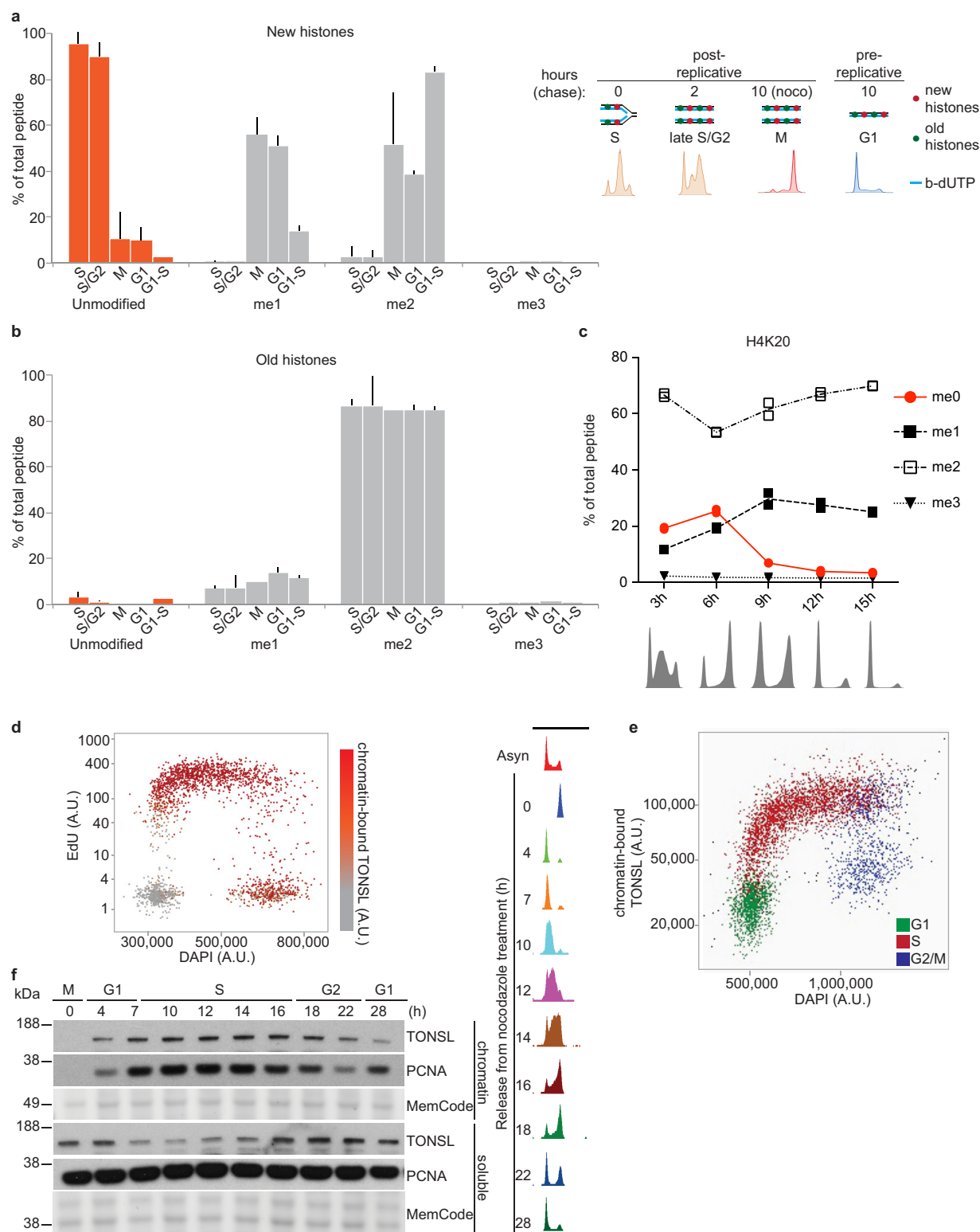
Extended Data Figure 4 | Structural comparison of the ARDs of TONSL and GLP. **a, b**, Representative view of the TONSL ARD with histone H4 tail (**a**; this work), and crystal structure of the GLP ARD in complex with histone H3 tail dimethylated at Lys9 (**b**; ref. 10). Both TONSL ARD and GLP ARD use the concave surface to bind their cognate target H4 tail and H3 tail, respectively. TONSL ARD recognizes H4K20me0 mainly

through three strong hydrogen bonds with acidic residues Glu530, Asp559 and Glu568, while GLP ARD recognizes H3K9me2 mainly through an aromatic cage forming by residues Trp839, Trp844, Glu847 and Trp877. **c**, ITC analysis of TONSL ARD binding to H3K9me1 peptide. **d**, ITC analysis of TONSL acidic stretch and ARD (amino acids 450–692) with H3K9me1 (amino acids 1–21) and H4 (amino acids 9–25) peptides.



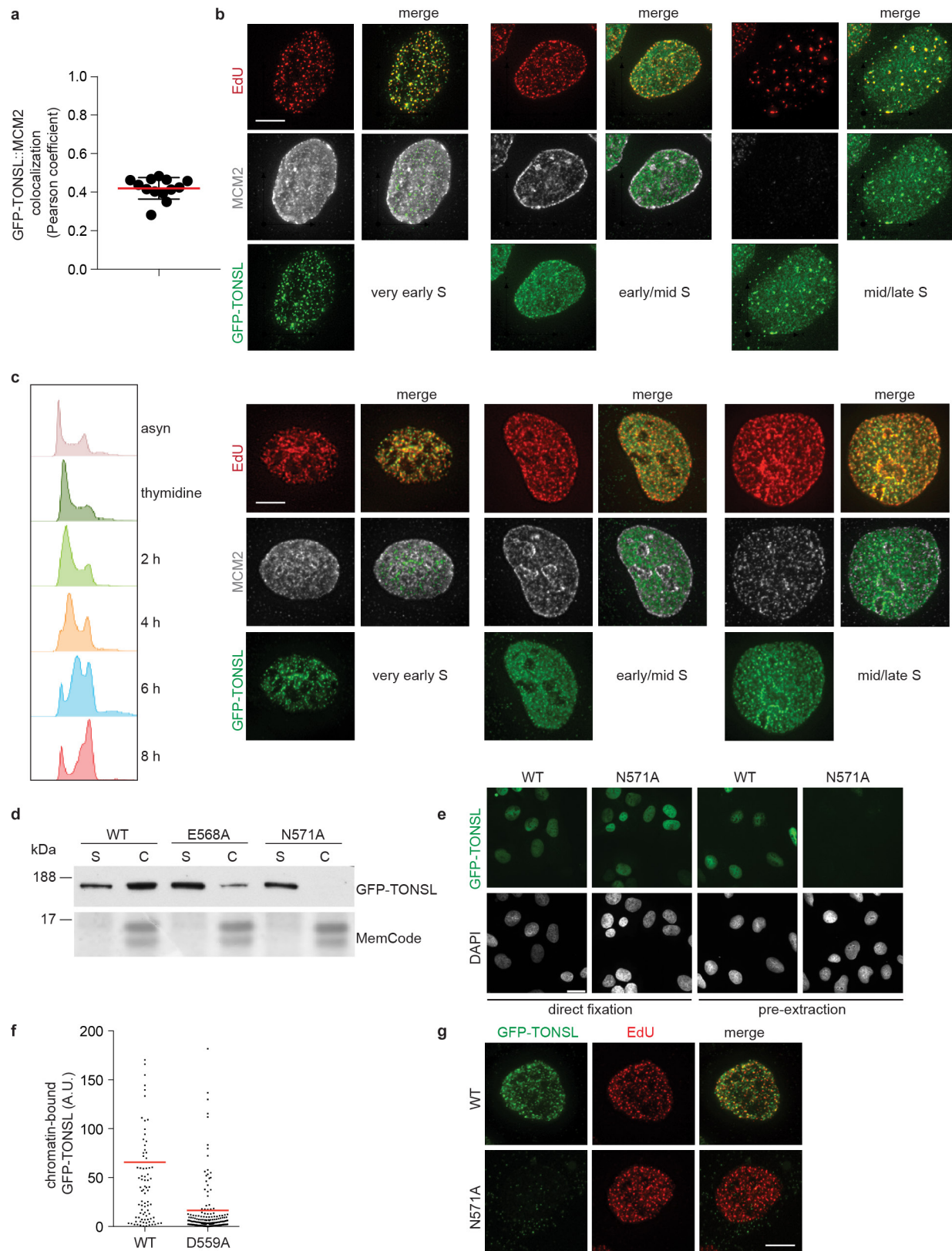
Extended Data Figure 5 | Effect of SET8 and MOF depletion on TONSL chromatin binding. **a**, Immunoprecipitation of GFP-TONSL from solubilized chromatin of GFP-TONSL U-2-OS cells (one representative experiment out of two is shown). Same exposures are shown for input and immunoprecipitation western blots of H3 and H4K16ac. **b**, TONSL ARD preference for H4K16ac could be mediated by I599 through hydrophobic association with the K16 acetyl group as I599E ARD mutation preferentially reduces binding to H4K16ac peptides as compared to the unmodified H4 tail. Left, pull-down of GFP-TONSL from cell extracts with biotinylated H4 tail peptides. Right, quantification of the western blot, GFP-TONSL binding to the H4K16ac peptide is shown relative to the unmodified peptide. Means with individual data points are shown ($n = 2$). **c**, High-content quantitative imaging of TONSL in pre-extracted U-2-OS cells. Plots show total chromatin-bound TONSL and DAPI intensities in cells treated with control or TONSL siRNA, confirming the specificity of

TONSL antibody staining. Each dot represents one nucleus. **d–f**, Analysis of TONSL chromatin-binding in MOF-depleted (**d**), SET8-depleted (**e**) and ionizing radiation (IR)-treated cells (**f**). Chromatin-bound TONSL was quantified by high content imaging of pre-extracted U-2-OS cells stained for endogenous TONSL. Mean TONSL intensity is shown. AU, arbitrary units. **d**, **e**, Knockdown efficiency and expected effect on histone modification were confirmed by western blotting (representative of two experiments). **e**, **f**, G1 cells were defined by gating on DAPI and EdU intensity. **f**, TONSL is not recruited to DNA damage in G1 cells, supporting that TONSL accumulation in SET8-depleted cells is due to lack of H4K20me1 and not DNA damage. Cells were irradiated (1.5 Gy) and analysed 1.5 h later (representative of two experiments). **d**, **f**, Error bars indicate s.d.; **d**, from left, $n = 4,920, 2,341, 3,608, 2,917$; **f**, $n = 382$ (–IR), 523 (+IR).



Extended Data Figure 6 | TONSL binding to chromatin during the cell cycle. **a, b**, H4K20 methylation levels on new and old histones analysed by NCC-pulse-SILAC (data are extracted from ref. 18). Cells grown in light SILAC medium were released into S phase in heavy medium and pulsed with b-dUTP. Chromatin was fixed, sonicated and b-dUTP-labelled fragments isolated on streptavidine beads by NCC. Histones were isolated and analysed by mass spectrometry for modifications on new (heavy) and old (light) histones. For clarity a 24 h (G1/S) chase time point is included. Error bars indicate s.d.; $n = 9$ (S), 3 (S/G2, M), 5 (G1), 3 (G1/S). Data for M (old histones) is shown as the mean of $n = 2$, as light peptides were not detected in one of the three biological replicates. **c**, H4K20 methylation levels measured by mass spectrometry in synchronized TIG3 fibroblasts.

d, Plot of mean EdU and total DAPI intensities from TIG3 fibroblasts as in Fig. 3b, with the intensity of chromatin-bound TONSL shown in the third dimension as a colour gradient. AU, arbitrary units. Each dot represents one nucleus. Note that a population of G2 cells (EdU negative) retain TONSL on chromatin. **e**, High-content quantitative imaging of pre-extracted U-2-OS cells stained for EdU and TONSL analysed as in Fig. 3b. **f**, Analysis of TONSL chromatin binding by cellular fractionation. U-2-OS cells released from a nocodazole block were followed by fluorescence-activated cell sorting (FACS) analysis of DNA content and analysed by western blotting of soluble (CSK-Triton extracted) and chromatin (pellet) fractions (representative of two experiments).

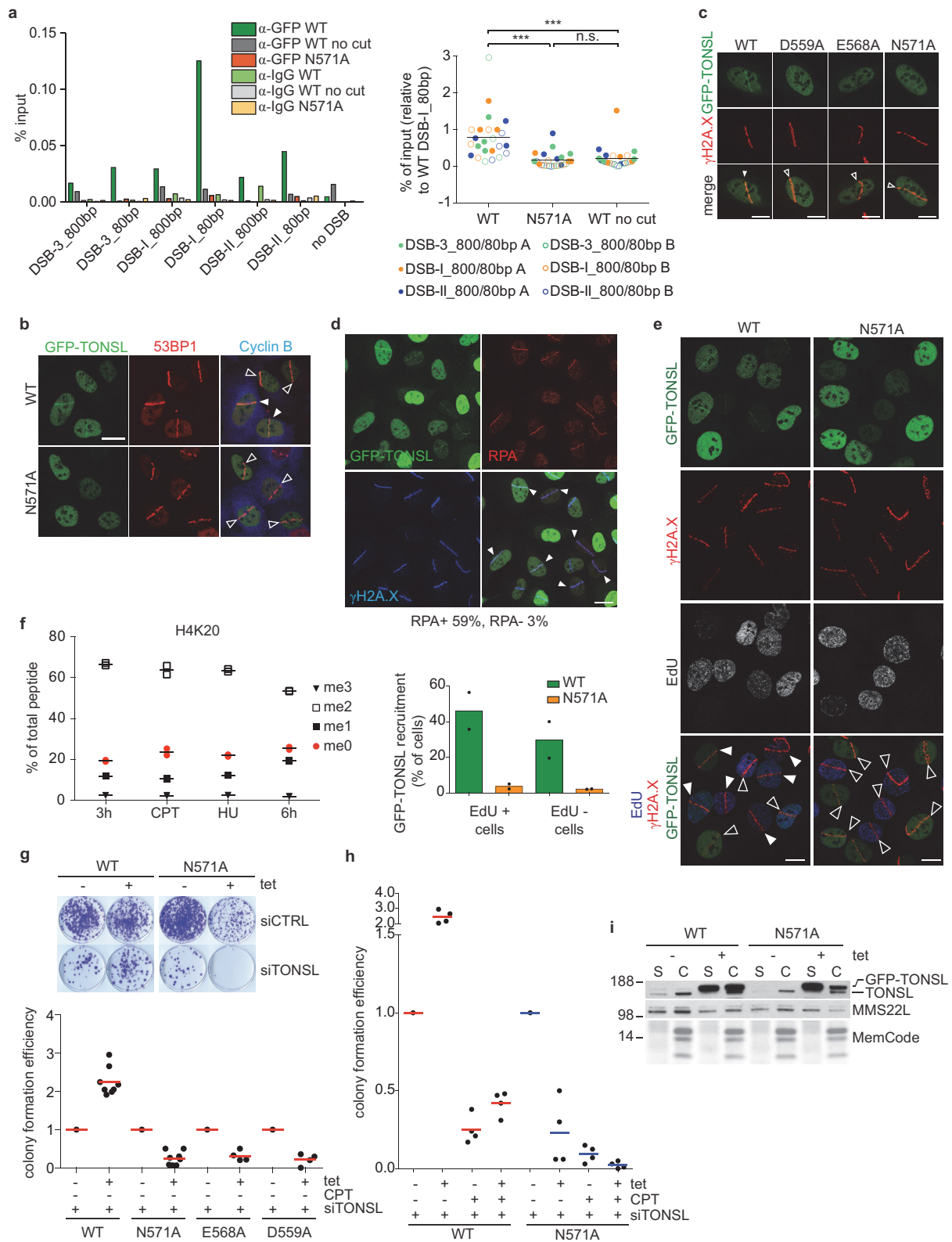


Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Analysis of GFP-TONSL localization.

a, Colocalization analysis of chromatin-bound GFP-TONSL with MCM2 analysed by deconvolution microscopy and measurement of Pearson coefficient in single cells. Error bars indicate s.d., $n = 13$ from two independent experiments. Representative image, Fig. 3c. **b, c**, Representative images for the analysis shown in Fig. 3d. Cells were either pulsed with EdU (40 μ M) for 15 min (**b**) or synchronized in G1/S and released into S phase in the continuous presence of EdU (5 μ M) (**c**). Images are representative of **b**: $n = 9$ (very early), 16 (early/mid), 10 (mid/late); **c**: 9 (very early), 27 (early/mid), 36 (mid/late). Scale bar, 5 μ m. **b**, EdU and MCM2 staining was used to determine the cell cycle state in asynchronous (asyn) cell populations. **c**, Progression through S phase

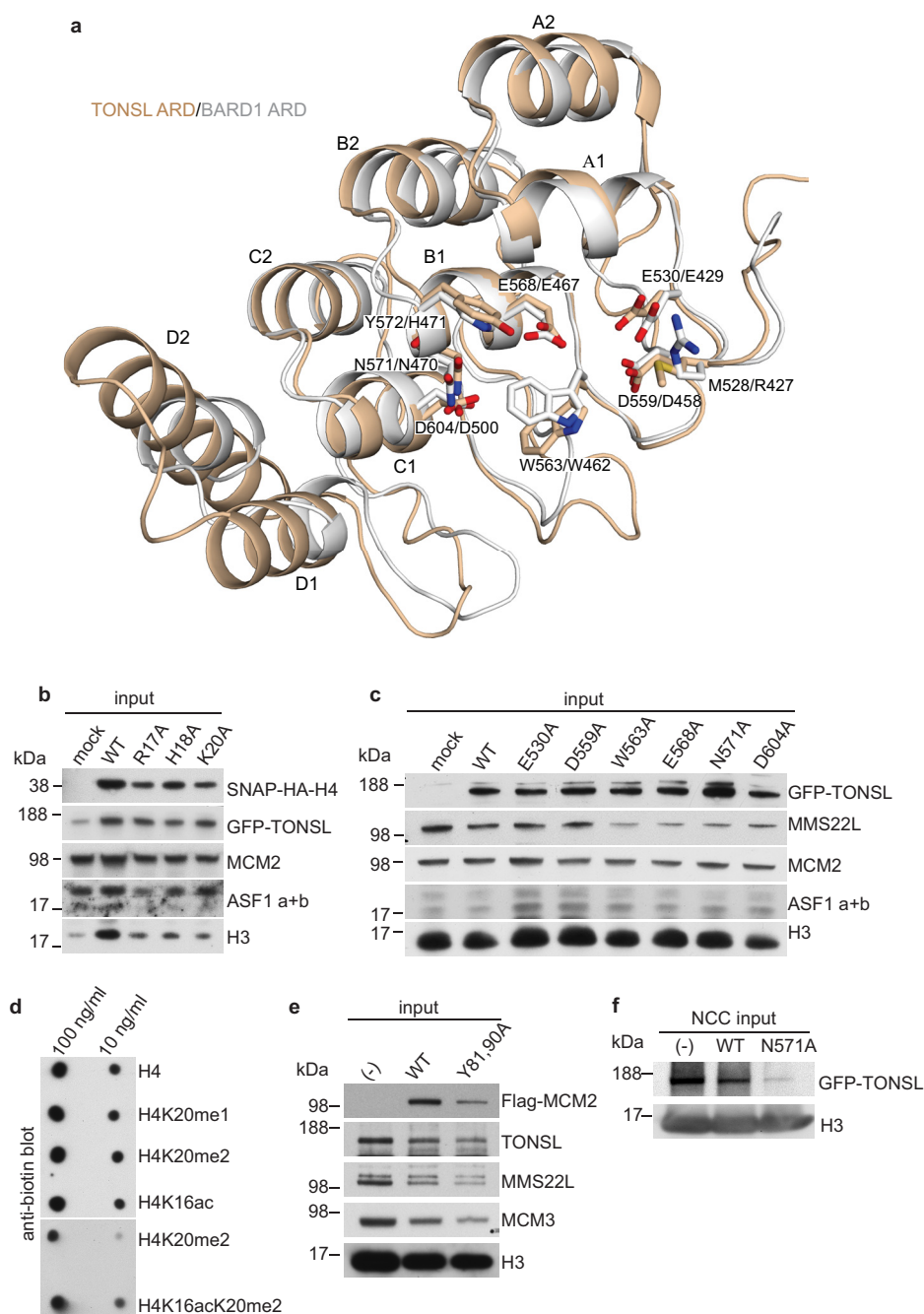
was followed by FACS analysis of DNA content. **d**, Chromatin-binding of GFP-TONSL analysed by cellular fractionation in inducible U-2-OS cells as quantified in Fig. 3e. C, chromatin; S, soluble. **e, f**, Chromatin-binding analysis as in Fig. 3f. U-2-OS cells conditional for GFP-TONSL ARD wild type (WT) and mutant were directly fixed or pre-extracted to remove soluble proteins. Data are representative of three (**e**) and two (**f**) experiments, fields of cells in **e** are representative of (from left) $n = 16$, 18, 17 and 17 images. Scale bar, 20 μ m. **g**, Asynchronous U-2-OS cells conditional for GFP-TONSL were pulsed with 40 μ M EdU for 15 min and soluble proteins were extracted. Representative images of EdU-positive cells are shown ($n = 30$ for wild type and N571A), for the specific patterns of TONSL wild type see Fig. 3d. Scale bar, 5 μ m.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | TONSL-MMS22L recruitment to damaged DNA. **a**, Left, ChIP-qPCR analysis of GFP-TONSL recruitment to site-specific DSBs induced by AsiSI, as shown in Fig. 4d but with additional controls. Note that the colours have been changed for clarity. Mean of technical duplicates is shown. Right, dot plot illustrating the relative enrichment of GFP-TONSL wild type (WT) and N571A obtained in four independent ChIPs performed on two biologically independent chromatin preparations. Each experiment was normalized to GFP-TONSL wild-type enrichment at DSB-I_80bp. Mean is shown with two-sided Mann-Whitney test; *** $P < 0.001$; not significant, $P > 0.05$; $n = 24$. Two-sided Mann-Whitney analysis of individual experiments gave similar results. **b**, U-2-OS cells conditional for GFP-TONSL were laser microirradiated. 53BP1 and cyclin B staining was used as markers of DNA damage and cells in S/G2 phase, respectively. Representative of three experiments as quantified Fig. 4e. Filled arrowheads indicate GFP-TONSL recruitment; open arrowheads indicate no recruitment. Scale bars, 10 μm . **c**, U-2-OS cells transiently transfected with GFP-TONSL wild type or the indicated mutants were laser microirradiated and processed for $\gamma\text{H2A.X}$ immunofluorescence. Representative cells are shown ($n = 200$ cells per condition from two independent experiments). **d**, U-2-OS cells conditional for GFP-TONSL were laser microirradiated. $\gamma\text{H2A.X}$ and RPA staining was used as markers of DNA damage and cells undergoing resection in S/G2 phase, respectively. The percentage of GFP-TONSL cells

with recruitment to RPA-positive (+) and RPA-negative (−) laser tracks is indicated. Data are representative of two independent experiments, a total of 118 cells were counted. **e**, Top, U-2-OS cells conditional for GFP-TONSL wild type and N571A were laser microirradiated. $\gamma\text{H2A.X}$ and EdU staining was used as markers of DNA damage and S phase cells, respectively. Bottom, quantification of GFP-TONSL cells with recruitment to laser tracks. Mean with individual data points are shown ($n = 2$, a total of 138 (wild type) and 174 (N571A) cells were counted). **f**, H4K20 methylation levels measured by mass spectrometry in synchronized TIG3 cells as in Extended Data Fig. 6c. Cells were released into S phase for 3 h and treated with hydroxyurea (HU; 3 mM) or CPT (1 μM) for 3 h or left untreated (6 h). Mean with individual data points are shown ($n = 2$). **g**, Colony formation in cells treated with control or TONSL siRNA and induced to express GFP-TONSL. As shown in Fig. 4f, but including additional mutants. Two cell concentrations in technical triplicate from two (E568A, D559A) or four (wild type, N571A) biological replicates are shown. **h**, Representation of the complementation analysis from Fig. 4f in a single panel including both CPT-treated and untreated cells. This illustrates that the toxicity of the TONSL ARD mutant is comparable to CPT treatment of cells expressing wild-type TONSL. **i**, Analysis of GFP-TONSL and MMS22L by cellular fractionation in cells inducible for GFP-TONSL ARD wild type and mutant. Representative experiment of the quantification shown in Fig. 4i.



Extended Data Figure 9 | Similarity of the ARDs in TONSL and BARD1, and protein inputs. **a**, Superposition of the structures of TONSL ARD and BARD1 ARD (Protein Data Bank accession 3C5R)²⁶. The main residues involved in TONSL ARD interactions with the H4 tail are compared to the corresponding residues of BARD1 ARD. The two ARDs show highly similar topology and conservation of the histone-binding surface. **b**, Input material of the experiment in Fig. 1h. **c**, Input material of the experiment in Fig. 1j. **d**, Spot assay with biotinylated H4 tail (amino acids 14–33) peptides confirming equal input into pull-down reactions.

e, Input material of the experiment in Fig. 4b. **f**, Input material of the NCC experiment in Fig. 4c. Note that because ARD mutation disrupts chromatin binding in the presence and absence of CPT (Figs 3e, f and 4a), GFP-TONSL N571 levels are low in the input chromatin. The NCC experiment in Fig. 4c supports our microscopy-based data (Fig. 4a) and further shows that there is no local accumulation of the GFP-TONSL ARD mutant at damaged forks that could have been missed in our microscopy-based quantification of total TONSL on chromatin.

Extended Data Table 1 | Data collection and refinement statistics

TONSL ARD - MCM2 HBD - H3/H4 Tetramer Complex	
Data collection	
Space group	P3 2 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	139.5, 139.5, 72.9
α , β , γ (°)	90, 90, 120
Resolution (Å)	50-2.43 (2.52-2.43)*
<i>R</i> _{pim} (%)	3.8 (46.8)
<i>I</i> / σ <i>I</i>	23.1 (1.8)
Completeness (%)	99.8 (99.7)
Redundancy	5.5 (5.5)
Refinement	
No. reflections (total/unique)	171,308/31,146
<i>R</i> _{work} / <i>R</i> _{free} (%)	20.1/24.6
No. atoms	
Protein	2,908
MES	12
GOL	12
Water	87
B-factors	
Protein	81.8
MES	108.5
GOL	92.6
Water	59.8
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.316

*Highest-resolution shell is shown in parenthesis. One crystal was used for the data.

Translation readthrough mitigation

Joshua A. Arribere¹, Elif S. Cenik¹, Nimit Jain², Gaelen T. Hess³, Cameron H. Lee³, Michael C. Bassik³ & Andrew Z. Fire^{1,3}

A fraction of ribosomes engaged in translation will fail to terminate when reaching a stop codon, yielding nascent proteins inappropriately extended on their C termini. Although such extended proteins can interfere with normal cellular processes, known mechanisms of translational surveillance¹ are insufficient to protect cells from potential dominant consequences. Here, through a combination of transgenics and CRISPR–Cas9 gene editing in *Caenorhabditis elegans*, we demonstrate a consistent ability of cells to block accumulation of C-terminal-extended proteins that result from failure to terminate at stop codons. Sequences encoded by the 3′ untranslated region (UTR) were sufficient to lower protein levels. Measurements of mRNA levels and translation suggested a co- or post-translational mechanism of action for these sequences in *C. elegans*. Similar mechanisms evidently operate in human cells, in which we observed a comparable tendency for translated human 3′ UTR sequences to reduce mature protein expression in tissue culture assays, including 3′ UTR sequences from the hypomorphic ‘Constant Spring’ haemoglobin stop codon variant. We suggest that 3′ UTRs may encode peptide sequences that destabilize the attached protein, providing mitigation of unwelcome and varied translation errors.

Failure to terminate translation at a stop codon can lead to ribosomes translating into a 3′ UTR. In some cases, translation may proceed through the 3′ UTR and into the poly(A) tail, triggering a process termed ‘nonstop decay’ and destabilizing both the mRNA and nascent protein (reviewed in ref. 1). However, for the majority of 3′ UTRs, a stop codon is encountered before the poly(A) tail^{2,3}. Readthrough events that encounter a subsequent termination codon are outside the scope of known translational surveillance pathways including nonstop¹. Depending on the 3′ UTR and the frame in which the ribosome enters, the late stop codon can be several, tens, or even hundreds of codons into a 3′ UTR, producing variant proteins with potentially problematic C-terminal appendages. This issue is highlighted by several pathologies caused by late frameshifts or stop-codon mutations in which 3′-UTR-encoded C-terminal extensions effect protein mislocalization^{4,5}, aggregation^{6,7}, and instability^{8–12}, with severe consequences for organisms. Depending on sequence, genetic background, conditions, and organism, estimates of readthrough efficiency vary from <1% to 10% or more, posing a potential problem of nontrivial magnitude^{10,13}.

We investigated whether, and to what extent, 3′ UTR translation has an effect on gene expression using a fluorescent reporter system in *C. elegans*. Initially, we selected 3′ UTRs from three genes: *unc-54* (encoding a muscle myosin), *tbb-2* (a beta tubulin), and *rpl-14* (a ribosomal protein). For each gene, fusion of the 3′ UTR to a green fluorescent protein (GFP)-driven by the *myo-3* promoter resulted in robust fluorescence in body-wall muscle (Fig. 1a). Next, by mutating stop codons, we created GFP reporters for each gene, which caused translation to read past the normal termination point, terminating instead at a stop codon part-way through the 3′ UTR (Fig. 1b). In each case, the ‘late stop’ reporter accumulated substantially less GFP, with differences in signal of at least tenfold. As a control, a co-injected mCherry marker was robustly expressed in the same cells. We conclude that translation

into the 3′ UTR can confer substantial loss of protein expression for at least these three 3′ UTRs in *C. elegans*.

To test whether translation into 3′ UTRs could confer a loss of protein expression more generally, a two-fluorescent-reporter system with each fluorophore transgene containing an identical 3′ UTR was used. Nine genes were chosen to reflect a variety of functions and expression levels: *rps-17* (small ribosomal subunit component), *r74.6* (*dom34/pelota* release factor homologue), *hlh-1* (muscle transcription factor), *eef-1A.1* (also known as *eft-3*, translation elongation factor), *myo-2* (a pharyngeal myosin), *mut-16* (involved in gene/transposon silencing), *bar-1* (a beta catenin), *daf-6* (involved in amphid morphogenesis), and *alr-1* (neuronal transcription factor). A criterion in choosing these genes was the presence (common for *C. elegans* genes, Extended Data Fig. 1) of an in-frame stop codon in the 3′ UTR at least 30 bases beyond the normal stop but upstream of known poly(A) sites. We fused the 3′ UTRs of each gene separately to GFP and mCherry, removing the canonical termination codon in the GFP construct. For each of the nine genes tested, the observed GFP signals were extremely faint, with raw GFP:mCherry fluorescence ratios of less than 0.1 (Fig. 1c, Extended Data Fig. 2). As a control, versions of the GFP reporter with the normal termination codon intact provided robust GFP expression, at least tenfold higher than the corresponding readthrough constructs (GFP:mCherry fluorescence ratios in the range of 0.3 to 0.9).

Several observations suggest how translation into 3′ UTRs might reduce protein levels. (1) Experiments with specific mutagenesis support a role for the eventual protein sequence. Shortening readthrough peptides (tested for *unc-54* and *tbb-2*) increased GFP expression (Fig. 2a). Extending this analysis, an equal-length non-synonymous substitution in the *unc-54* 3′ UTR restored GFP expression, whereas synonymous substitution with multiple base differences did not. (2) Mutagenesis analysis of constructs using a constant 3′ UTR reinforced the inference of peptide sequence as the primary determinant of GFP loss. We found that the nucleotide sequence between the normal termination codon and the first in-frame termination codon was sufficient to confer GFP loss if inserted at the end of the GFP coding region for *unc-54*, *tbb-2*, *hlh-1*, *daf-6*, *rps-20*, or *rps-30* (Fig. 2b). The *rps-30* readthrough region had the weakest effect on GFP, and was the shortest (nine amino acids). We performed further mechanistic dissection by synonymous variation of readthrough regions from *unc-54*, *tbb-2*, and *rps-20*, with GC contents from 35–60%, in some cases mutating >50% of bases. Each synonymously substituted variant conferred robust loss of GFP. (3) Decreased expression following translation into the 3′ UTR required peptide linkage between the upstream protein and the 3′-UTR-encoded segment. To assess the relationship between covalent linkage with the translated C-terminal peptide and the outcome for the larger protein, we took advantage of a picornavirus-derived oligopeptide sequence that causes cleavage and release of the nascent chain, after which ribosomes continue translation of the downstream sequence^{14,15}. Insertion of the T2A peptide (EGRGSLTTCGDVEENPGP) between GFP and the *unc-54* 3′-UTR-encoded sequence rescued GFP expression, whereas an uncleavable T2A* point mutant did not (Fig. 2b). Restoration of GFP levels by T2A to the level of no-insert controls

¹Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. ²Department of Bioengineering, Stanford University, Stanford, California 94305, USA.

³Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA.

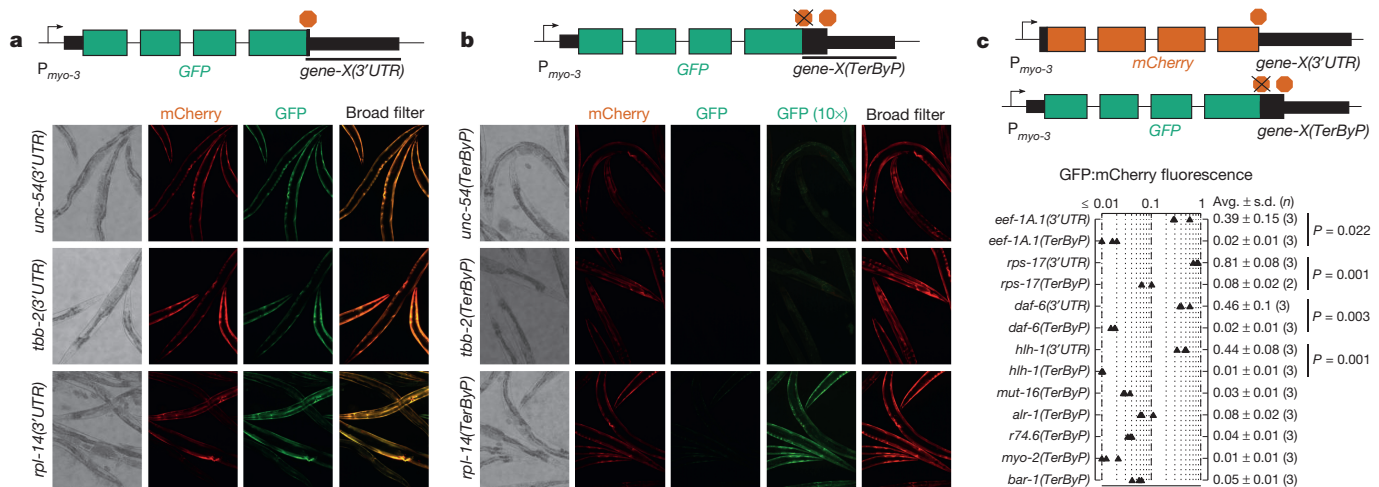


Figure 1 | Translation into 3' UTRs results in substantial loss of protein expression. **a**, Dual fluorescence reporter assay to test expression with different 3' UTRs. Transgenic arrays of each GFP construct were created using *pha-1* selection and mCherry (pCFJ104) as a coinjection marker. Broad filter detects GFP and mCherry signals simultaneously; deviation from yellow towards red or green shows more mCherry or GFP fluorescence, respectively. Three independent transgenic lines were made for each (two for *tbb-2(TerByP)*); transgenic lines with similar mCherry expression are shown. 200 ms exposure, 10 × objective. **b**, Dual fluorescence reporter assay to test expression of readthrough for different 3' UTRs. The stop codon of each 3' UTR was mutated, allowing translation to proceed into the 3' UTR. TerByP refers to 'Termination ByPass', the

region between the canonical termination codon and first in-frame termination codon in the 3' UTR. Images were collected as in **a**. 'GFP (10×)' is a 2 s exposure. The dim yellowish fluorescence in 'GFP (10×)' for *unc-54(TerByP)* and *tbb-2(TerByP)* is autofluorescence. **c**, For each gene, the 3' UTR was fused to mCherry and GFP. GFP expression was tested with the stop codon mutated to a sense codon (*TerByP*). For each of *eef-1A.1*, *rps-17*, *daf-6*, and *hlh-1*, GFP expression was also tested with the normal stop codon in place (3' UTR). The ratio of GFP to mCherry fluorescence under a broad fluorescence filter was used as a metric (Extended Data Fig. 2, Methods). Each triangle represents an independently generated transgenic line; mean and s.d. of *n* (shown in parentheses) lines shown. Student's *t*-test two-tailed *P* value.

also argues against mRNA destabilization as a substantial factor in the protein loss observed upon readthrough.

The above results could be explained if GFP was generally incompatible with C-terminal fusions in our system. To address this, we inserted a variety of sequences downstream of GFP: 3 × Flag, 3 × haemagglutinin (HA), three random sequences created *in silico*, and six arbitrary fragments of in-frame coding sequence from *C. elegans*

genes, approximately length-matched to 3'-UTR-encoded sequences (Fig. 2b). GFP expression varied between constructs but was generally higher than 3'-UTR-encoded sequences: 3 × HA, 3 × Flag, 2 out of 3 random sequences, and 4 out of 6 coding-derived fragments exhibited GFP:mCherry fluorescence ratios of >0.13, higher than all nine tested 3'-UTR-derived C-terminal extensions and significant statistically (*P* = 0.004, Kolmogorov–Smirnov test). Thus the effects of

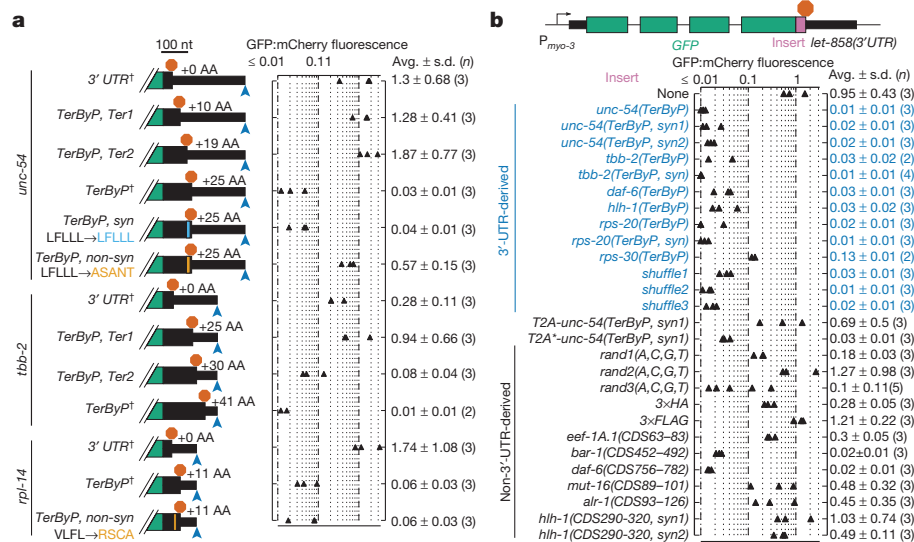


Figure 2 | Identification of determinants for product loss upon translation into the 3' UTR. **a**, Shortening or non-synonymous mutations of the readthrough region can restore GFP expression. Stop codons and/or mutations were inserted into each GFP::3'-UTR fusion as shown, with stop codons (red stop sign) and poly(A) site (blue arrowhead). † indicates the constructs shown in Fig. 1. mCherry (pCFJ104) was used as a coinjection marker. '+X AA' indicates amino acids added relative to cognate control ('+0 AA') construct. Constructs and mutated regions drawn to scale, scale bar at top. Mean and s.d. of *n* (shown in parentheses) lines shown.

b, 3'-UTR-encoded peptides are sufficient to confer GFP loss. Sequences were inserted upstream of the *let-858* 3' UTR. 'syn', synonymously substituted variants. *Shuffle1–3* contain shuffled codons of *unc-54*, *tbb-2*, and *rpl-14* (VFLF to RSCA) TerByP regions (Extended Data Fig. 4). T2A, 'self-cleaving' peptide which releases the upstream nascent chain; T2A*, a non-cleaving variant^{14,15}. *Rand1–3* (A, C, G, T) are random combinations of A, C, G, and T created *in silico*. CDS N–M is an arbitrary fragment of the respective coding DNA sequence of the gene (from amino acid N to M).

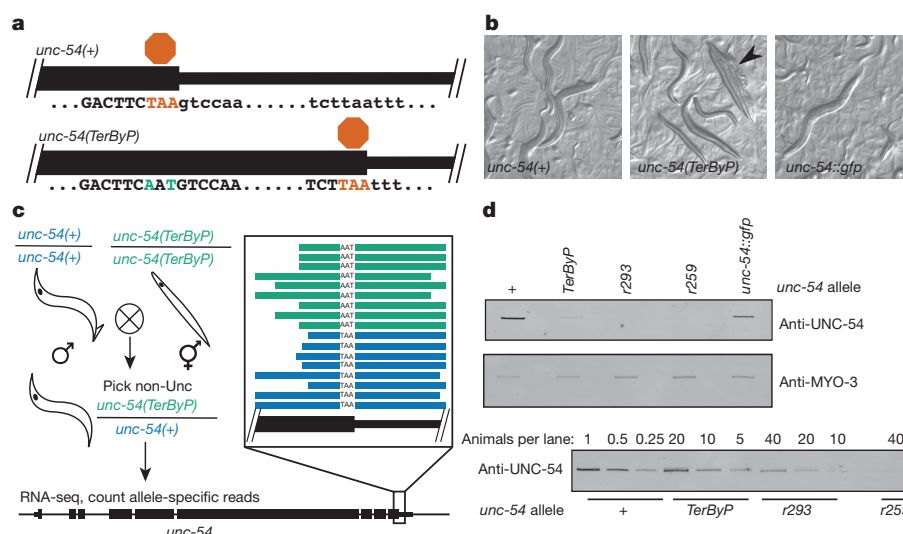


Figure 3 | Translation into the 3' UTR at an endogenous locus decreases protein levels. **a**, Schematic of wild-type and readthrough alleles of *unc-54*, the latter made using CRISPR–Cas9 genome editing¹⁷. See Extended Data Table 1 for additional loci and edits. **b**, Brightfield images of *unc-54* alleles. Arrowhead indicates a 'bag of worms', the shell of an egg-laying-defective mother consumed by its retained progeny. **c**, RNA-seq from *unc-54(TerByP/+)* heterozygotes showed no differential effect on RNA levels. *unc-54(TerByP/+)* heterozygotes were chosen among progeny of *unc-54(+)* males crossed with *unc-54(TerByP)* homozygotes and

allele-specific reads identified. The framed inset shows individual allele-specific RNA-seq reads (bars) from *unc-54(TerByP)* (AAT, green) and *unc-54(+)* (TAA, blue). See also Extended Data Figs 5, 6. **d**, Quantification of UNC-54 protein levels. Immunoblotting was performed on homozygous populations of the indicated animals. *unc-54(r293)* encodes a nonsense-mediated decay allele of *unc-54*, producing <5% of normal UNC-54 protein. *unc-54(r259)* contains a >17 kb deletion spanning most of the *unc-54* locus. For the lower blot, the number of animals loaded per lane is indicated. For gel source data, see Supplementary Fig. 1.

3'-UTR-encoded sequences are not explained by a general intolerance of GFP to C-terminal extensions (see also Methods, Extended Data Figs 3, 4).

It was conceivable that peculiarities of GFP and/or transgene expression systems might underlie the above observations. To establish effects of 3' UTR translation at endogenous genes, we sought loci where (1) a loss of protein would be detectable phenotypically, (2) C-terminal fusions are known to be functional, (3) the next in-frame stop codon of the endogenous locus is ≥ 10 amino acids past the annotated stop codon, yet upstream of annotated poly(A) sites¹⁶, and (4) there is little or no autoregulation/feedback. *unc-54* and *unc-22* meet all of the criteria, and *pha-4*, *unc-45*, and *tra-2* at least the first three points (Methods). For each locus, we mutated the stop codon to allow translation into the 3' UTR¹⁷ (Fig. 3a). In parallel, we analysed small insertions/deletions generating late frameshifts for *unc-22* and *unc-54*. Additional controls had length-matched sequences and/or GFP tags at the C terminus (Extended Data Table 1). For each of *unc-22*, *unc-45*, *unc-54*, and *tra-2*, translation into the 3' UTR in at least one frame generated a strong hypomorphic (near null) phenotype specific to each locus. Other C-terminal tags for each gene did not cause loss of expression, although one *tra-2* C-terminal tag did produce a Tra phenotype. The ability to place alternative tags on the C terminus without obvious phenotypic consequences argues against a general sensitivity of the C terminus to tagging. For *unc-22* and *unc-54*, that elongation into the 3' UTR in only some frames elicited a hypomorphic phenotype suggests that ribosome elongation into the 3' UTR is not detrimental *per se*.

To determine the consequences on gene expression upon translation into 3' UTRs, we analysed the effects of the *unc-54(cc3389)* TAA(stop) to AAT(Asn) mutation on RNA, translation, and protein output. We analysed mRNA expression in *unc-54(cc3389/+)* heterozygotes (phenotypically wild type to avoid complications from an Unc phenotype). RNA-seq revealed that the *unc-54(cc3389)* and wild-type alleles were at approximately equal amounts in the mRNA pool, suggesting that 3' UTR translation does not appreciably destabilize the *unc-54* mRNA (Fig. 3c). In parallel, we detected an ~ 20 -fold reduction in UNC-54 protein in immunoblots in *unc-54(cc3389)* mutants (Fig. 3d). To look for possible alterations in translation for *unc-54(cc3389)*, we examined the distribution of RNase-protected mRNA fragments

with ribosome footprint profiling¹⁸. We observed no significant difference in the loading of ribosomes on *unc-54* mRNA (Extended Data Fig. 5), nor on the number, distribution, frame, or fragment size of ribosomes in the extended region (Extended Data Fig. 6).

A model that arises from these observations is that 3'-UTR-encoded peptides mark their resulting products for destruction, either co- or post-translationally. Conceivably this process might operate either in a specific cell/tissue type or in a broad spectrum of different contexts. A broadly-expressed reporter bearing a readthrough extension would be expected to highlight any tissue that failed to destabilize the C-terminal peptide. Using a broadly expressed promoter (*unc-37*) driving GFP with and without the *unc-54* 3'-UTR-encoded peptide, we observed no cells where GFP was robustly retained (data not shown).

We likewise considered the possibility that 3'-UTR-encoded peptides might act to limit protein levels in human cells, developing a specific assay using a lentiviral dual fluorescence reporter encoding puromycin *N*-acetyl-transferase tethered to mCherry-T2A, followed by eGFP and a multiple cloning site (Fig. 4a). The resulting reporter expresses both fluorophores from the same mRNA, yet as two disjoint polypeptides, allowing consideration of the effect of a peptide tag on eGFP expression independent of effects on mCherry and mRNA expression. We validated the split dual fluorophore approach in K562 cells using tags known to be destabilizing (d1ODC, d4ODC¹⁹) or not (3 \times Flag, 3 \times HA) (Fig. 4b). We selected 13 genes of varying expression and function, and inserted the region between the annotated termination codon and first-in-frame termination codon downstream of eGFP. For 9 of 13 genes, the readthrough region reduced the eGFP:mCherry fluorescence ratio between 3- and 30-fold, a stronger reduction than the degenon d4ODC (Fig. 4c). Although not universal, the substantial loss of eGFP fluorescence for a majority of readthrough regions opens up the possibility that translation into 3' UTRs may be generally inhibitory to expression across systems.

We hypothesize that a function of 3' UTRs is to minimize the accumulation of extended protein products that could be produced through translational readthrough.

This feature may prove generally important in causes of genetic disease. For example, readthrough alleles (such as stop to Gln) of the *HBA2* locus in humans produce a fraction ($\sim 1\%$) of normal HBA2

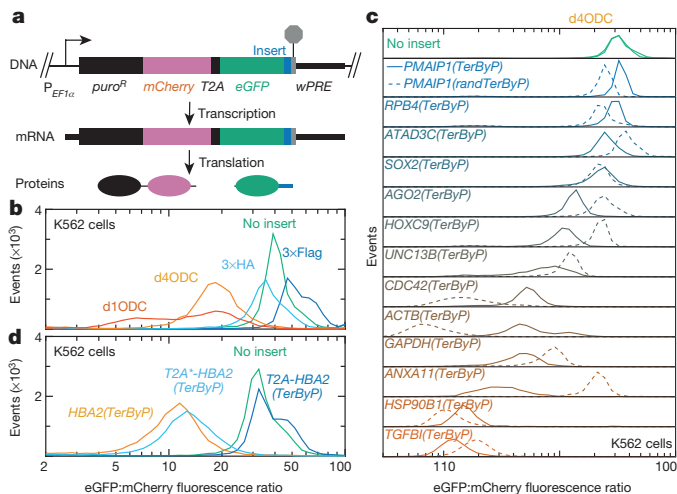


Figure 4 | Translation into 3' UTRs results in protein loss for several genes in humans. **a**, Lentiviral reporter schematic. A puroR–mCherry fusion was co-translationally cleaved from eGFP–insert by T2A. Constructs in **b–d** were integrated into K562 cells via lentiviral infection and puromycin selection. **b**, Validation of dual fluorescence reporter. Inserts downstream of eGFP were 3× Flag, 3× HA, and degrons d4ODC (half-life, ~4 h¹⁹), d1ODC (half-life, ~1 h¹⁹). **c**, The sequence between the annotated and first in-frame termination codon (TerByP) from each gene was inserted downstream of eGFP (solid line). For comparison, nucleotides of each TerByP region were randomized, producing a length- and nucleotide-frequency-matched construct (randTerByP, dashed line). Cells with eGFP lacking an insert and grown a week apart (top, green solid lines) and approximate fluorescence ratio of d4ODC (orange line) are shown. **d**, The first 30 amino acids of the HBA2 3' UTR were inserted downstream of eGFP (orange). Insertion of a self-cleaving T2A peptide restored expression (blue), whereas an uncleavable mutant (T2A*) did not (light blue).

protein (α -globin), causing thalassemia. Translation into the HBA2 3' UTR is known to destabilize the HBA2 mRNA²⁰, but it is unclear what effect the appended C-terminal 31 amino acids have on HBA2 protein. We considered the possibility that the HBA2 3' UTR-encoded peptide might prevent protein expression in humans, contributing to the loss of HBA2 protein. When appended to eGFP, the HBA2 3' UTR-encoded peptide decreased the eGFP:mCherry fluorescence ratio in K562 cells (Fig. 4d). Furthermore, eGFP fluorescence was rescued by a self-cleaving (but not an uncleavable mutant) T2A peptide.

Several observations from the literature support the notion that 3' UTR-encoded peptides may be detrimental to expression for more genes and organisms than those assessed here. In *Saccharomyces cerevisiae*, translation past a point in the HIS3 3' UTR confers a substantial loss in protein expression, without detectable effects on mRNA levels¹¹. Similarly, readthrough of the cyclic AMP phosphodiesterase PDE2 stop codon produces a destabilized protein variant, and this has been suggested to explain elevated cyclic AMP levels in PSI⁺ yeast¹⁰. Differential stability by polymorphisms in the readthrough peptide of SKY1 has been postulated to explain [PSI]-induced strain differences in diamide sensitivity²¹. Particularly intriguing are recent findings that stop codon mutations at the c-FLIP_L locus confer protein instability for this anti-apoptotic factor in mice, leading to embryonic lethality¹². The same study also noted several hereditary human disease alleles where 3' UTR-encoded peptides are destabilizing, conferring marked decreases in protein activity and level⁸.

Not every case of stop codon readthrough is destabilizing^{4–7} (Fig. 4c), and some readthrough events are functional and regulated to defined levels^{22–25}. Understanding the mechanisms by which some readthrough events are detected and cleared (whereas others are not) may prove informative in biological contexts and pathological states where inappropriate readthrough occurs. We do not yet know the determinants of a

translated 3' UTR sequence that confer loss of protein, though the ability of numerous sequences (including shuffled and randomized 3' UTR variants, Fig. 2b, 4c) suggests that a highly degenerate sequence is sufficient. Consistent with the idea that the effects of readthrough peptides may be mediated via their biophysical characteristics, we observed a significant negative relationship between hydrophobicity and expression of GFP (K562 cells, *C. elegans*) and endogenous loci (*unc-22* and *unc-54*, *C. elegans*) (Extended Data Figs 7–9, Supplementary Information).

Destabilization by 3' UTR-encoded peptides could effectively mitigate at least three types of events in which a stop codon is inappropriately bypassed: (1) stop codon misreading (for example, by suppressor tRNAs). Suppressor tRNAs permit readthrough of up to 30% of ribosomes at a stop codon (UAA, UAG, or UGA)¹³. Whereas some suppressor tRNAs can be toxic, other cells tolerate even high levels of readthrough^{13,26–28}. Destabilization of readthrough products by C-terminal appendages may effectively buffer cells from suppressor tRNA-induced proteostatic chaos. (2) A ribosomal frameshift in a coding region which is late enough that no premature termination codon is encountered. In this case, ribosomes would enter the 3' UTR out-of-frame with the coding region. In our manipulations, translation of 3' UTRs in multiple frames was detrimental to expression (Extended Data Table 1, data not shown), and similar amino acid and hydrophobicity biases hold for all three 3' UTR frames (Extended Data Figs 8, 9). (3) Errors in RNA processing or ribosome dysfunction could produce a variety of other improperly terminated peptides from which translation readthrough mitigation would provide valuable relief.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 October 2015; accepted 5 May 2016.

Published online 1 June 2016.

- Klauer, A. A. & van Hoof, A. Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdiscip. Rev. RNA* **3**, 649–660 (2012).
- Hamby, S. E., Thomas, N. S., Cooper, D. N. & Chuzhanova, N. A meta-analysis of single base-pair substitutions in translational termination codons ('nonstop' mutations) that cause human inherited disease. *Hum. Genomics* **5**, 241–264 (2011).
- Williams, I., Richardson, J., Starkey, A. & Stansfield, I. Genome-wide prediction of stop codon readthrough during translation in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **32**, 6605–6616 (2004).
- Falini, B. et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* **352**, 254–266 (2005).
- Hollingsworth, T. J. & Gross, A. K. The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *J. Biol. Chem.* **288**, 29047–29055 (2013).
- Vidal, R. et al. A stop-codon mutation in the BRI gene associated with familial British dementia. *Nature* **399**, 776–781 (1999).
- Vidal, R. et al. A decamer duplication in the 3' region of the BRI gene originates an amyloid peptide that is associated with dementia in a Danish kindred. *Proc. Natl Acad. Sci. USA* **97**, 4920–4925 (2000).
- Pang, S. et al. A novel nonstop mutation in the stop codon and a novel missense mutation in the type II 3 β -hydroxysteroid dehydrogenase (3 β -HSD) gene causing, respectively, nonclassic and classic 3 β -HSD deficiency congenital adrenal hyperplasia. *J. Clin. Endocrinol. Metab.* **87**, 2556–2563 (2002).
- Clegg, J. B., Weatherall, D. J. & Milner, P. F. Haemoglobin Constant Spring—a chain termination mutant? *Nature* **234**, 337–340 (1971).
- Namy, O., Duchateau-Nguyen, G. & Roussel, J. P. Translational readthrough of the PDE2 stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **43**, 641–652 (2002).
- Inada, T. & Aiba, H. Translation of aberrant mRNAs lacking a termination codon or with a shortened 3' UTR is repressed after initiation in yeast. *EMBO J.* **24**, 1584–1595 (2005).
- Shibata, N. et al. Degradation of stop codon read-through mutant proteins via the ubiquitin-proteasome system causes hereditary disorders. *J. Biol. Chem.* **290**, 28428–28437 (2015).
- Capone, J. P., Sharp, P. A. & RajBhandary, U. L. Amber, ochre and opal suppressor tRNA genes derived from a human serine tRNA gene. *EMBO J.* **4**, 213–221 (1985).
- Ahier, A. & Jarriault, S. Simultaneous expression of multiple proteins under a single promoter in *Caenorhabditis elegans* via a versatile 2A-based toolkit. *Genetics* **196**, 605–613 (2014).
- Doronina, V. A. et al. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol. Cell. Biol.* **28**, 4227–4239 (2008).

16. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* **469**, 97–101 (2011).
17. Arribere, J. A. *et al.* Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**, 837–846 (2014).
18. Ingolia, N. T., Ghaemmamghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
19. Yen, H.-C. S., Xu, Q., Chou, D. M., Zhao, Z. & Elledge, S. J. Global protein stability profiling in mammalian cells. *Science* **322**, 918–923 (2008).
20. Liebhaber, S. A. & Kan, Y. W. Differentiation of the mRNA transcripts originating from the alpha 1- and alpha 2-globin loci in normals and alpha-thalassemics. *J. Clin. Invest.* **68**, 439–446 (1981).
21. Torabi, N. & Kruglyak, L. Genetic basis of hidden phenotypic variation revealed by increased translational readthrough in yeast. *PLoS Genet.* **8**, e1002546 (2012).
22. Steneberg, P. & Samakovlis, C. A novel stop codon readthrough mechanism produces functional Headcase protein in *Drosophila* trachea. *EMBO Rep.* **2**, 593–597 (2001).
23. Freitag, J., Ast, J. & Bölker, M. Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. *Nature* **485**, 522–525 (2012).
24. Eswarappa, S. M. *et al.* Programmed translational readthrough generates antiangiogenic VEGF-Ax. *Cell* **157**, 1605–1618 (2014).
25. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
26. Waterston, R. H. A second informational suppressor, *sup-7 X*, in *Caenorhabditis elegans*. *Genetics* **97**, 307–325 (1981).
27. Laski, F. A., Ganguly, S., Sharp, P. A., RajBhandary, U. L. & Rubin, G. M. Construction, stable transformation, and function of an amber suppressor tRNA gene in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **86**, 6696–6698 (1989).
28. Hudziak, R. M., Laski, F. A., RajBhandary, U. L., Sharp, P. A. & Capecchi, M. R. Establishment of mammalian cell lines containing multiple nonsense mutations and functional suppressor tRNA genes. *Cell* **31**, 137–146 (1982).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Fire Laboratory for critical reading of the manuscript, C. Frøkjær-Jensen and K. Artiles for technical expertise, and T. Schedl, T. Inada, C. Joazeiro, L. Ling, A. Nager, and N. Spies for discussions. A. Sapiro and B. Li were instrumental in developing the RNA-seq2 protocol. This work was supported by grants from NIH R01GM37706, T32HG000044 (G.T.H.), 1DP2HD084069-01 (M.C.B.), 5F32GM112474-02 (J.A.A.), Walter and Idun Berry Foundation (E.S.C.), and NSF DGE-114747 (C.H.L.).

Author Contributions J.A.A., E.S.C., and A.Z.F. designed *C. elegans* experiments. J.A.A. and E.S.C. conducted *C. elegans* experiments. N.J. developed the RNA-seq2 protocol. J.A.A. performed computational analyses. J.A.A. conducted experiments in human cell lines, as designed and aided by J.A.A., G.T.H., C.H.L., M.C.B., and A.Z.F. J.A.A. and A.Z.F. wrote the paper with help from all authors.

Author Information Sequencing data are available at Sequence Read Archive (SRP064516). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Z.F. (afire@stanford.edu).

Reviewer Information *Nature* thanks J. S. Butler, M. Yarus and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

C. elegans strain construction and husbandry. *C. elegans* were grown at 23 °C on agar plates with nematode growth medium seeded with *Escherichia coli* strain OP50 as described²⁹. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). A full list of strains used is available in Supplementary Table 1.

Transgenic array-containing strains were generated as follows: PD5102 (*pha-1(e2123ts); rde-1(ne300)V*) young adult hermaphrodites (grown at 16 °C) were injected with a mix of 90 ng μl^{-1} pC1 (containing a rescuing fragment of *pha-1*), 5 ng μl^{-1} of an mCherry-containing vector, and 5 ng μl^{-1} of a GFP-containing vector. Unless otherwise indicated, GFP was driven by the *myo-3* promoter to drive expression in the body-wall muscle³⁰. Injectants were shifted to 23 °C to select for F1 progeny animals bearing a transgenic array (selecting for *pha-1(+)* expression³¹). The *rde-1* allele included in this strain avoided a modest degree of secondary siRNA-based silencing observed with many extrachromosomal transgenes³². For transgenic lines generating low levels of GFP, we considered the possibility that the GFP protein was toxic and selected against. Under this model, one might expect (1) a subset of sick and/or dead GFP-positive F1 animals, (2) muscle defects due to muscle-specific expression of potentially-toxic GFP derivatives, (3) concomitant low levels of mCherry, and/or (4) a decrease in the efficiency with which transgenic lines were obtained³². None of these effects were observed, arguing against any contribution of negative selection to the observed low GFP expression.

For a subset of strains, we deviated from the above protocol to generate *pha-1* arrays as follows: (1) whereas most transgenic lines were generated from independently injected parents, a handful of strains were possibly generated from siblings of an injected parent (PD6480, 6481, 6482, 6483, 6484, 6485, 6486, 6493, 6494, 6495). In these cases, all injectants were pooled together on the same plate, and independent F1 were picked off to generate transgenic lines. Previous work has demonstrated independent F1 from the same injected parent carry distinct transgenic arrays^{32,33}. (2) During the course of our analyses, we found some strains with an mCherry-negative subpopulation. The mCherry-positive subpopulation was isolated and propagated to generate the strains PD6401, 6450, 6452, 6456, 6457, and 6464.

CRISPR-Cas9 genome editing was performed in the VC2010 (PD1074) N2 background as described¹⁷. We selected *pha-4* (refs 34, 35), *unc-45* (refs 36, 37), *tra-2* (refs 38–40), *unc-22* (ref. 29), and *unc-54* (refs 26, 38, 41, 42) based on the criteria in the text (citations indicated). The statement that *unc-22* and *unc-54* exhibit little or no autoregulation/feedback is based on a number of genetic experiments (with heterozygous²⁹, amber-suppressed²⁶, and/or *smg*-suppressed³⁸ alleles) which express either UNC-54 or UNC-22 at stable intermediate levels (between wild type and null). Alleles of *unc-45* were initially generated in the VC2010 background, though the embryonic lethality made *unc-45(TerByP)* difficult to maintain. We subsequently remade all alleles in a balanced heterozygote background (*sCI(s2023)(dpy-1(s2170)) III(+)*) and considered non-Dpy segregants for phenotypic analyses.

Human cell line construction. K562 cells (obtained from ATCC) were grown at a density of ~ 0.5 to 1×10^6 cells ml^{-1} in RPMI medium supplemented with penicillin/streptomycin, L-glutamine, and 10% FBS. All cell lines were maintained in a humidified incubator (37 °C, 5% CO₂), and checked regularly for mycoplasma contamination. As a means of validating K562 cells, we performed RNA-seq on a subset of lines and observed good correlation with published data sets⁴³ (data not shown). Viral particles were produced in HEK293T cells in 6-well dishes, and 1 ml of viral supernatant was used to infect $\sim 100,000$ K562 cells by spin infection, 10^3 relative centrifugal force for 2 h. Polybrene was omitted to keep the infection rate low ($< 10\%$), ensuring a single incorporation event for most cells. After 3 days of recovery, cells were selected with puromycin at $0.7 \mu\text{g ml}^{-1}$ for at least 3 days. Fluorescence was examined on a BD Accuri C6 flow cytometer, with appropriate gating for live cell events and investigators blinded to cell line identity. For each construct examined via puromycin selection in K562 cells, similar eGFP and mCherry fluorescence levels were also observed in transient transfection in HEK293T cells in the absence of puromycin, arguing against a puromycin-selected skew in mCherry fluorescence.

Plasmids. Plasmids were constructed by restriction digest or Gibson cloning as detailed in Supplementary Table 2. pJA138/L3785 and pJA137/pCFJ104 were used as the basis of all *C. elegans* GFP- or mCherry-containing vectors, respectively. Portions of pMCB306 and pMCB309 were used to construct pJA291, the parental *puro::mCherry::T2A::eGFP::MCS::wPRE* vector for experiments in human cells. Plasmids were confirmed by both sequencing and restriction digest, and plasmid concentrations determined with the Qubit dsDNA Broad Range kit (Invitrogen). Plasmids that may be useful have been deposited with Addgene: pJA327 (*C. elegans* superfolder GFP in L3785), pJA291, pJA317 (pJA291 with d1ODC insert) and pJA318 (pJA291 with d4ODC insert).

GFP fusions were constructed with a GFP variant that corresponds to wild-type (*Aequora*) GFP with mutations at position 65 (Ser to Thr for human, Ser to Cys for *C. elegans*) known to improve folding and acquisition of fluorescence. Even with these mutations, GFP has a known propensity to misfold under some circumstances, and we therefore examined the effect of a subset of the 3'-UTR-encoded sequences (*hlh-1*, *daf-6*, and *unc-54*) downstream of a faster and more robust-folding GFP variant, superfolder GFP (ref. 44). The observed reduction in superfolder-GFP:mCherry ratios was quantitatively similar to that observed with normal GFP (Extended Data Fig. 3).

Sequences of Flag⁴⁵, HA⁴⁶, d1ODC¹⁹, and d4ODC¹⁹ were obtained from the indicated publications. For exact sequences, see Supplementary Tables 1 and 2. T2A was previously shown to function in *C. elegans*¹⁴. Translation elongation through a member of the 2A peptide family (consensus D(V/I)EXNPGP) causes ribosomal pause, then release of the N-terminal peptide up to and including the Gly¹⁵. Translation elongation resumes, with the C-terminal peptide being produced with an N-terminal Pro.

Microscopy. Animals were immobilized by placing on a slide with a coverslip in 5 mM EDTA, 50 mM NaCl, 1 mM levamisole and imaged on a Nikon Eclipse E6000 microscope using a Nikon super high pressure mercury lamp power supply. Filter cubes for fluorescence images were GFP (96342, Nikon Corp), mCherry (96321, Nikon Corp), and broad (GFP and mCherry, 59022, Chroma Technology Corp). Images were collected with a 3CCD Digital Camera C7780 (Hamamatsu Corp) using HCLImage (Version 1.0.2.060107, Hamamatsu Corp). Images of PD4251 and one of PD3363/3364 were taken for each imaging session and compared to ensure consistency between days.

For quantification of GFP to mCherry relative fluorescence, animals were imaged using a $4 \times$ objective with a broad filter and 200 ms exposure. Investigators were blinded during imaging. To avoid image over- or underexposure, a number of exceptionally bright or dim strains were taken with a decreased or increased (respectively) exposure time (PD1798, 40 ms; PD3294, 500 ms; PD3299, 50 ms; PD3395, 500 ms; PD6327, 50 ms; PD6375, 50 ms; PD1786, 50 ms; PD1789, 50 ms; PD1790, 50 ms; PD6460, 500 ms; PD6469, 50 ms; PD6471, 50 ms; PD6472, 50 ms; PD6473, 50 ms; PD6485, 100 ms; PD1787, 40 ms; PD6450, 100 ms; PD6455, 40 ms; PD6477, 40 ms; PD6479, 50 ms; PD6498, 40 ms; PD6499, 40 ms; PD6500, 40 ms; PD6501, 40 ms; PD6502, 40 ms; PD6503, 40 ms; PD6504, 40 ms).

Raw pixel values for the red and green channels were obtained from image files using the tiffle package in python. Pixels below a threshold distance (200) from the median pixel intensity of the entire image were discarded as background. Pixels above a threshold intensity distance (4,000 of a possible 4,095) from the origin were discarded as saturated. The median pixel intensity for the entire image (essentially the black background, given the relatively low density of *C. elegans* tissue) was subtracted from the remaining pixels, and the slope of the linear regression line taken as the GFP:mCherry fluorescence ratio. This metric was robust to different exposure times and neutral density filters.

Statistics. Statistical tests and *P* values are stated throughout the text and figures.

To test statistical significance of C-terminal appendage effects on the GFP:mCherry fluorescence ratio (Fig. 2b), we divided the data into two groups: (1) 3'-UTR-derived (*unc-54(TerByP)*, *tbb-2(TerByP)*, *daf-6(TerByP)*, *hlh-1(TerByP)*, *rps-20(TerByP)*, *rps-30(TerByP)*, *shuffle1-3*); and (2) non-3'-UTR-derived (*rand1-3(A,C,G,T)*, $3 \times \text{HA}$, $3 \times \text{Flag}$, *eef-1A.1(CDS63-83)*, *bar-1(CDS452-492)*, *daf-6(CDS756-782)*, *mut-16(CDS89-101)*, *alr-1(CDS93-126)*, *hlh-1(CDS290-320,syn1)*). For each construct, we took the average GFP:mCherry fluorescence ratio of all available lines. We compared the distribution of 3'-UTR-derived and non-3'-UTR-derived GFP:mCherry fluorescence ratio values by Kolmogorov-Smirnov test ($P = 0.004$).

No statistical methods were used to predetermine sample size.

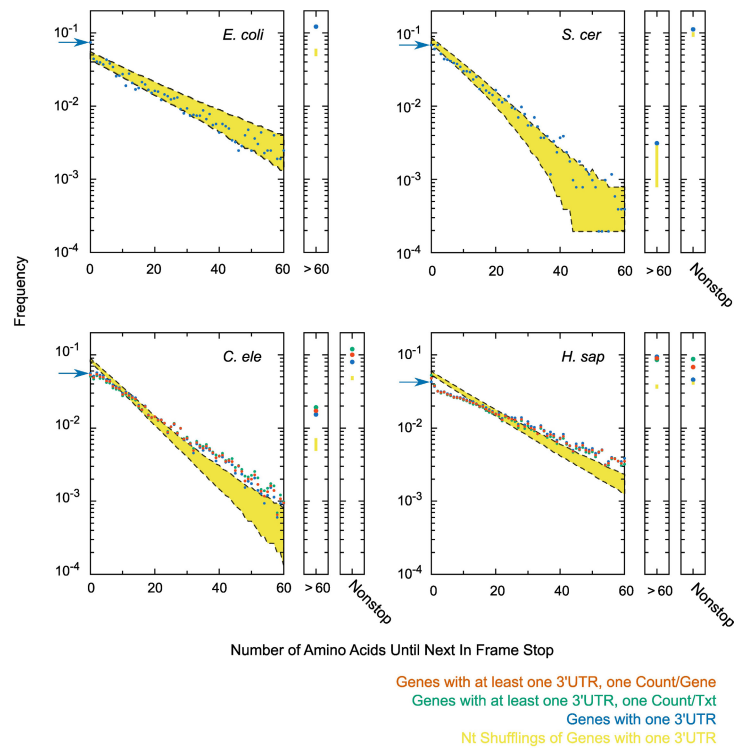
Ribosome footprint profiling. Ribo-seq was performed essentially as previously described^{18,47}, with a few modifications. Briefly, animals were grown to around L4 stage, and collected by centrifugation and flash-freezing in liquid nitrogen. Animals were ground with a mortar and pestle in liquid nitrogen, after which the powder was thawed in excess volume ice-cold polysome lysis buffer (20 mM Tris (pH 8.0), 140 mM KCl, 1.5 mM MgCl₂, 1% Triton) with cycloheximide ($100 \mu\text{g ml}^{-1}$). RNase 1 and sucrose gradient centrifugation was performed as previously described⁴⁷. Around 2 μg of purified, RNase-1-digested monosomal RNA was run on a urea 15% polyacrylamide gel, and the entire region from ~ 15 –30 nt was excised for library preparation. At this point, the protocol continued with T4 polynucleotide kinase (PNK) (New England Biolabs) treatment as with the RNA-seq1 protocol (next section).

RNA sequencing. Two RNA sequencing (RNA-seq) protocols were used in this study. The first RNA-seq protocol (RNA-seq1) was performed on homozygote populations of animals (Extended Data Fig. 5). 5 μg of total RNA was treated with the RiboZero kit (Illumina). RNA was fragmented at 95 °C for 30 min by addition of an equal volume of 100 mM sodium carbonate, 0.5 mM EDTA (pH 9.3) buffer. RNA

fragments were gel-purified, then treated with T4 PNK. 3'-ligation with AF-JA-34.2 adaptor (/5rApp/NNNNNAGATCGGAAGAGCACACGTCT/3ddC/, Integrated DNA Technologies) and T4 RNA ligase 1 (New England Biolabs) was performed at room temperature for 4 h with 20% PEG8000 in 3.3 mM DTT, 8.3 mM glycerol, 50 mM HEPES KOH (pH 8.3), 10 mM MgCl₂, 10 µg ml⁻¹ acetylated BSA. Unligated AF-JA-34.2 was removed by sequential treatment with 5'deadenylase (M0331S, New England Biolabs), then RecJ_F (M0264S, NEB). Reverse transcription was carried out with AF-JA-126 (/5Phos/AGATCGGAAGAGCGTCGTGT/iSp18/CACTCA/iSp18/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, Integrated DNA Technologies) as a primer. Circular ligase treatment and PCR were as previously described⁴⁷.

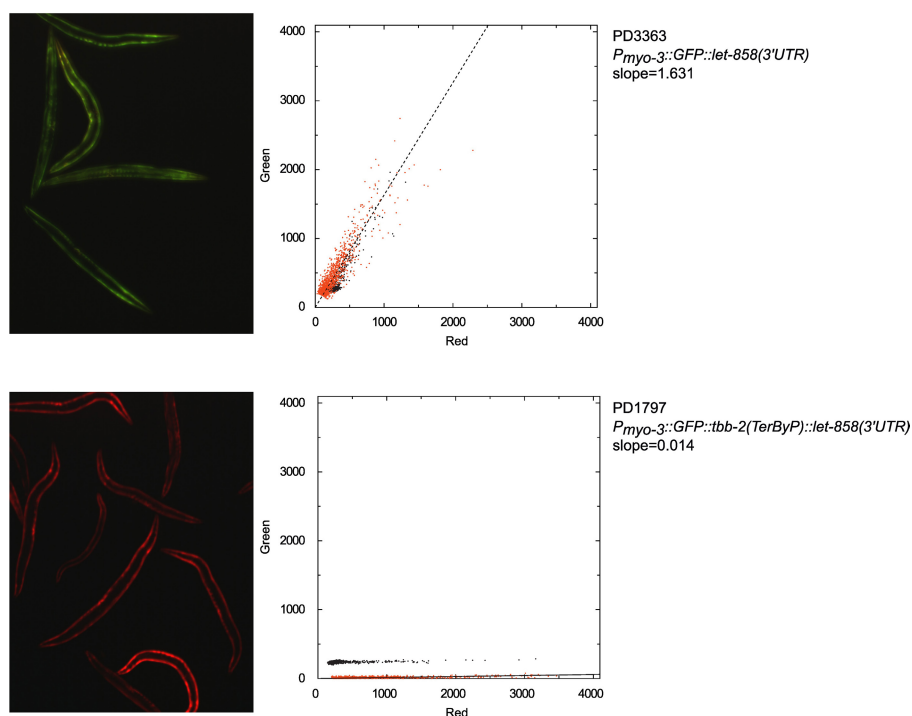
A second RNA-seq protocol (RNA-seq2) was used to examine RNA levels with small numbers of heterozygote animals (Fig. 3C). Around sixty L2–L4 mixed gender animals were picked and flash-frozen in 50 mM NaCl, and RNA extracted with trizol. RNaseH and 94 oligonucleotides complementary to ribosomal RNA were used to deplete rRNA from the sample⁴⁸. Briefly, ~250 ng of a cocktail of DNA oligonucleotides complementary to rRNA (Supplementary Table 3, ordered from Integrated DNA Technologies) was mixed with ~100 ng total RNA in 125 mM Tris (pH 7.4), 250 mM NaCl in 8 µl. The sample was heat-denatured at 95 °C for 2 min, then cooled at -0.1 °C per s to 45 °C. 1 µl of digestion buffer was added (500 mM Tris (pH 7.4), 1 M NaCl, 200 mM MgCl₂) with 1 µl (5 units) Thermostable RNase H (Epicentre), and the sample was incubated at 45 °C for one hour. DNA oligonucleotides were removed by treatment with TURBO DNase (ThermoFisher) at 37 °C, and RNA was extracted using an equal volume of phenol/chloroform. An RNA-seq library was prepared using the SMARTer Stranded RNA-Seq kit (Clontech Laboratories, Inc.). **Sequencing.** Libraries were sequenced on a MiSeq Genome Analyzer (Illumina, Inc.). Reads were mapped to the *C. elegans* genome (Ensembl70, WBcel215) using STAR (v2.3.1 ref. 49), with the mutated bases of *unc-54(cc3389)* and *unc-54(e1301)* masked. For Ribo-seq and RNA-seq1, reads bearing the same last 6 nucleotides (from NNNNNN, added with AF-JA-34.2) were assumed to be PCR duplicates and collapsed to a single read. For RNA-seq2, multiple reads containing the same start and stop mapping positions were collapsed to a single read count to reduce effects of PCR bias. The removal of PCR duplicates with either protocol only affected ~5–10% of reads and did not adversely impact any of the analyses shown. RNA-seq1 and Ribo-seq were performed once for each strain shown in Extended Data Figs 5, 6. **Genomes and annotations.** Although we sought to use the latest genome versions and annotations, we found it prudent to take advantage of the care and time with which other researchers annotated and analysed earlier versions of genomes. For whole genome alignment of nematode species, *C. elegans* UCSC genome ce10/WS220 was used. To examine the length of predicted C-terminal extensions upon readthrough (Extended Data Fig. 1), genomes and annotations of each of the indicated species were as follows: *E. coli* Ensembl genome and annotations from assembly GCA_000967155.1.30, *S. cerevisiae* genome S288c (R57-1-1_20071212) and annotations⁵⁰, *C. elegans* UCSC genome (WS190/ce6) and annotations¹⁶, *H. sapiens* Ensembl genome release 83 and annotations from TargetScan v7.0⁵¹. **Immunoblotting.** Animals were boiled in 1 × SDS loading buffer (65 mM Tris (pH 6.8), 10% glycerol, 2% SDS, 2 mM PMSE, 1 × Halt Protease Inhibitor (Thermo), 10% 2-mercaptoethanol) and run on a 7.5% Criterion TGX gel (Bio-Rad Laboratories, Inc.). Protein was transferred to a low background fluorescence PVDF membrane (Millipore). The membrane was blocked in 3% nonfat milk in 1 × PBST with 250 mM NaCl. The 5-6 antibody was used at a 1:5000 dilution to detect *myo-3*, and 5-8 antibody used at a 1:5000 dilution to detect *unc-54* (ref. 52). The 5-6 and 5-8 monoclonal antibodies were produced previously by purification of endogenous myosin proteins. Secondary antibody staining was performed with 1:500 Cy3-conjugated affiniPure goat anti-mouse (Jackson ImmunoResearch). Imaging was performed on a Typhoon Trio (Amersham Biosciences), and quantification performed in ImageJ. For the lower blot of Fig. 3d, lysates were made from multiple animals, and serial dilutions performed to titrate the number of animals per lane.

29. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
30. Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. & Fire, A. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* **135**, 385–404 (1993).
31. Granato, M., Schnabel, H. & Schnabel, R. *pha-1*, a selectable marker for gene transfer in *C. elegans*. *Nucleic Acids Res.* **22**, 1762–1763 (1994).
32. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
33. Stinchcomb, D. T., Shaw, J. E., Carr, S. H. & Hirsh, D. Extrachromosomal DNA transformation of *Caenorhabditis elegans*. *Mol. Cell. Biol.* **5**, 3484–3496 (1985).
34. Mango, S. E., Lambie, E. J. & Kimble, J. The *pha-4* gene is required to generate the pharyngeal primordium of *Caenorhabditis elegans*. *Development* **120**, 3019–3031 (1994).
35. Zhong, M. *et al.* Genome-wide identification of binding sites defines distinct functions for *Caenorhabditis elegans* PHA-4/FOXA in development and environmental response. *PLoS Genet.* **6**, e1000848 (2010).
36. Venolia, L. & Waterston, R. H. The *unc-45* gene of *Caenorhabditis elegans* is an essential muscle-affecting gene with maternal expression. *Genetics* **126**, 345–353 (1990).
37. Ao, W. & Pilgrim, D. *Caenorhabditis elegans* UNC-45 is a component of muscle thick filaments and colocalizes with myosin heavy chain B, but not myosin heavy chain A. *J. Cell Biol.* **148**, 375–384 (2000).
38. Hodgkin, J., Papp, A., Pulak, R., Ambros, V. & Anderson, P. A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* **123**, 301–313 (1989).
39. Hodgkin, J. A. & Brenner, S. Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. *Genetics* **86**, 275–287 (1977).
40. Mapes, J., Chen, J.-T., Yu, J.-S. & Xue, D. Somatic sex determination in *Caenorhabditis elegans* is modulated by SUP-26 repression of *tra-2* translation. *Proc. Natl Acad. Sci. USA* **107**, 18022–18027 (2010).
41. Anderson, P. & Brenner, S. A selection for myosin heavy chain mutants in the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **81**, 4470–4474 (1984).
42. Eide, D. & Anderson, P. The gene structures of spontaneous mutations affecting a *Caenorhabditis elegans* myosin heavy chain gene. *Genetics* **109**, 67–79 (1985).
43. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* **306**, 636–640 (2004).
44. Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
45. Hopp, T. P. *et al.* A short polypeptide marker sequence useful for recombinant protein identification and purification. *Nat. Biotechnol.* **6**, 1204–1210 (1988).
46. Field, J. *et al.* Purification of a RAS-responsive adenyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* **8**, 2159–2165 (1988).
47. Stadler, M. & Fire, A. Wobble base-pairing slows *in vivo* translation elongation in metazoans. *RNA* **17**, 2063–2073 (2011).
48. Morlan, J. D., Qu, K. & Sinicropi, D. V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* **7**, e42882 (2012).
49. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).
51. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, 1–38 (2015).
52. Miller, D. M. III, Ortiz, I., Berliner, G. C. & Epstein, H. F. Differential localization of two myosins within nematode thick filaments. *Cell* **34**, 477–490 (1983).
53. Thompson, O. *et al.* The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. *Genome Res.* **23**, 1749–1762 (2013).
54. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).



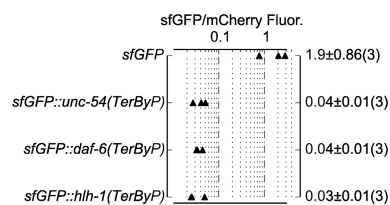
Extended Data Figure 1 | Distribution of C-terminal extensions upon stop codon readthrough. Annotations and genomes were as described in Supplementary Methods. Each 3' UTR was translated starting one codon after the stop codon until the next in-frame stop codon. For metazoans, counting was performed in three different ways: including only genes for which exactly one 3' UTR was annotated (blue), counting each annotated 3' UTR separately (green), or counting each gene once and splitting gene counts with multiple 3' UTRs equally amongst the 3' UTR isoforms (red). 'Nonstop' denotes 3' UTRs for which no stop codon was encountered before the poly(A) tail. For each species the distribution of next in-frame stop codons was calculated for 1,000 nucleotide shuffling of 3' UTR sequences for genes with a single 3' UTR annotated, and 95% confidence interval shown (yellow). A similar 'randomized' distribution was obtained

upon shuffling 3' UTR sequences and preserving dinucleotide frequency. The frequency of stops immediately after the annotated stop codon (amino acid length 0) is highlighted with a blue arrow in each species. The distribution of peptide lengths follows an exponential decay curve, where the slope is related to the probability of encountering a stop codon at each position. In the simplest model, the probability of encountering a stop codon is constant throughout the 3' UTR, accounting for the roughly linear shape of each plot (previously noted^{2,3}). Notable exceptions are a tendency towards second in-frame stops in *E. coli* (blue arrow), and a tendency towards peptides >60 amino acids in length in all species. In *E. coli*, the enrichment towards longer downstream peptides is at least partially explained by the operonic layout of genes.

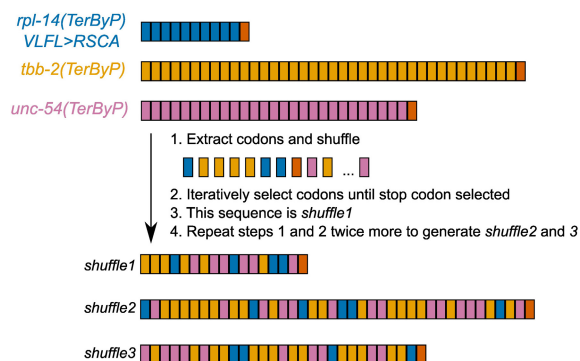


Extended Data Figure 2 | Example quantification of the GFP:mCherry fluorescence ratios of images. Images were taken under a broad excitation and emission filter to allow for simultaneous capture of GFP and mCherry fluorescence. Intensities of each pixel in the red and green channels were extracted in python. Unfiltered pixel intensities are shown as black

dots. Pixels were filtered, background subtracted, and linear regression performed (red dots and line, see Methods). For simplicity, the green-red intensities from 1,000 random pixels are shown. The GFP:mCherry fluorescence ratio was taken as the slope of the linear regression line. $10\times$ objective.

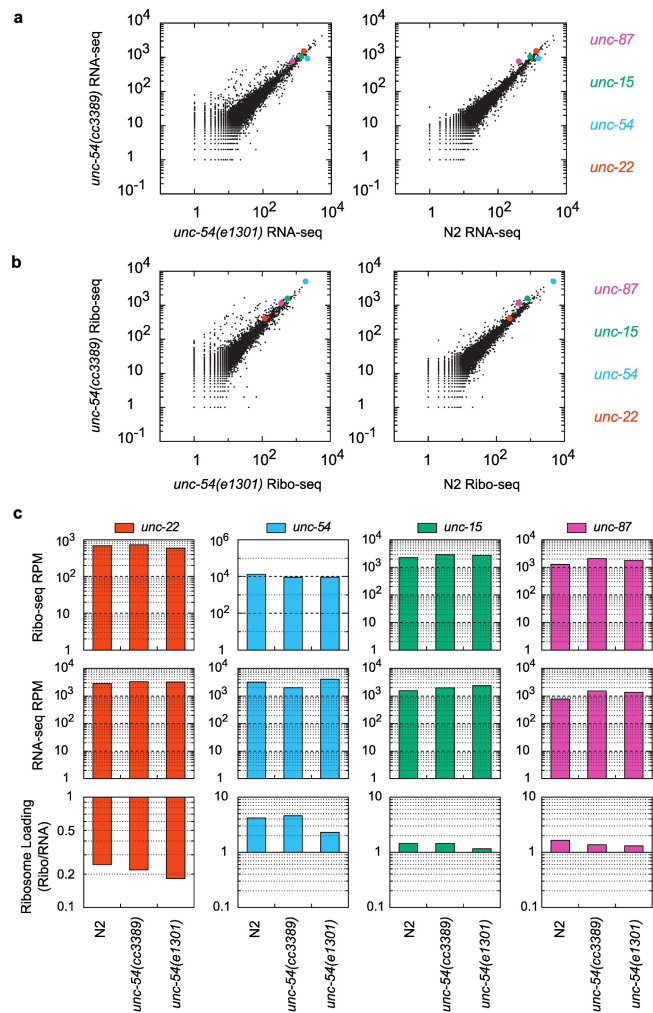


Extended Data Figure 3 | Readthrough regions confer a loss of superfolder GFP fluorescence. Each of the indicated TerByP regions were inserted downstream of superfolder (sf) GFP, upstream of the *let-858* 3' UTR. TerByP is the region after the annotated stop codon, up to and including the first in-frame stop codon in the 3' UTR. Quantification was performed as described in Extended Data Fig. 2.

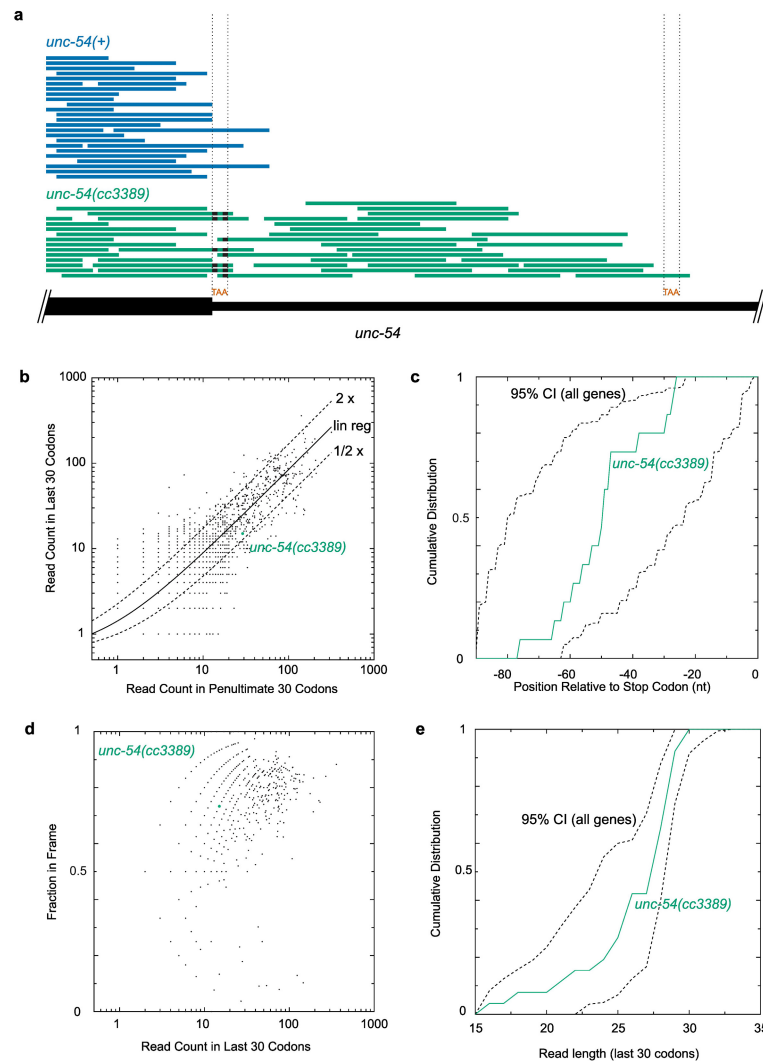


Extended Data Figure 4 | Explanation of 'shuffle' sequences.

Trinucleotide codons from each TerByP region are colour-coded by gene (top). Codons were extracted and randomly shuffled in python. A codon was iteratively selected until a stop codon was encountered, defining *shuffle1*. The process was repeated twice more to define *shuffle2* and *shuffle3*. The resulting shuffle peptides are a combination of all three TerByP regions. Lengths and colour-coding of codons for *shuffle1*–3 accurately reflect the sequences they are derived from.



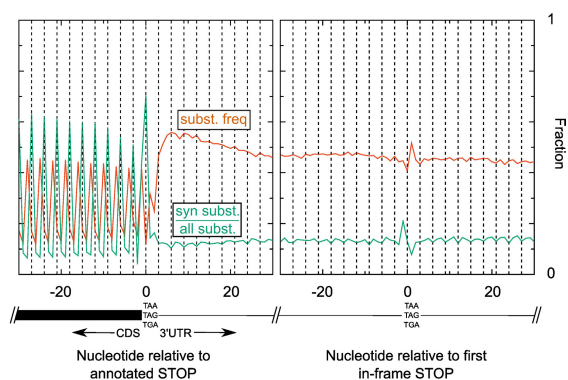
Extended Data Figure 5 | RNA-seq and ribo-seq from *unc-54* mutants. **a–c**, RNA-seq (**a**) and ribosome footprint profiling (ribo-seq) (**b**) library mRNA counts, with summary counts (**c**) for the indicated strains and mRNAs. Libraries were prepared from L4 animals, as described Methods. ‘N2’ is wild type (PD1074, VC2010 (ref. 53)). *unc-54(cc3389)* bears a TAA (stop) to AAT (Asn) mutation, *unc-54(TerByP)*. *unc-54(e1301)* has a GGA (Gly387) to AGA (Arg387) point mutation that confers a temperature-sensitive Unc phenotype with minimal discernible effects on UNC-54 protein levels. *unc-54(e1301)* was included as a control for the Unc phenotype of *unc-54(cc3389)*, though *e1301* confers a less severe Unc phenotype than *cc3389*. Values for *unc-54* mRNA (blue) are highlighted throughout, and for comparison, three additional transcripts known to be at least partly expressed in the body-wall muscles are also highlighted: *unc-87* (pink), *unc-15* (green), and *unc-22* (red).



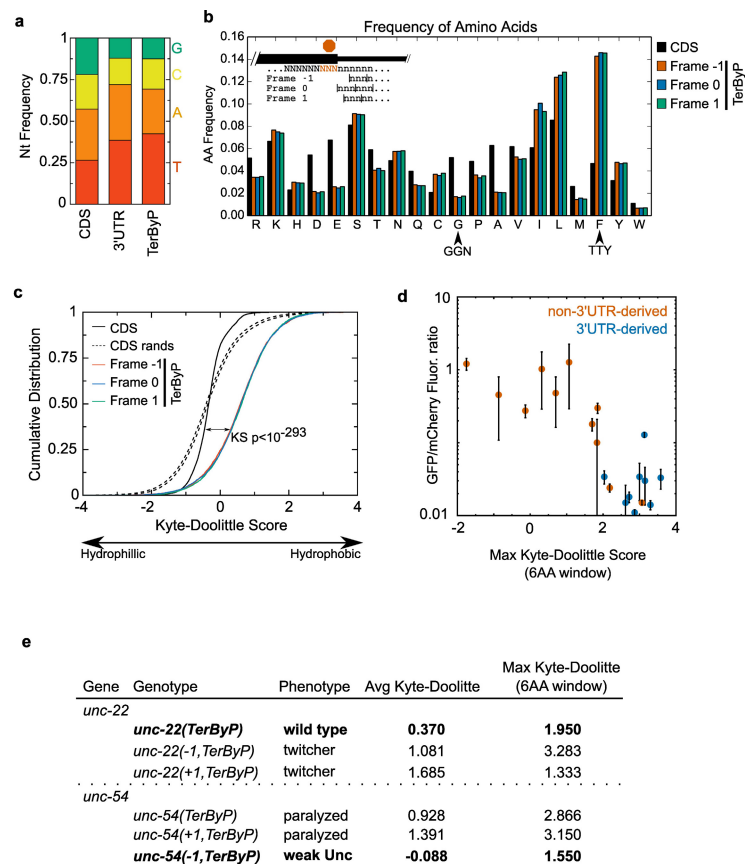
Extended Data Figure 6 | Ribo-seq of *unc-54(cc3389)* shows an unexceptional progression of ribosomes in the readthrough region.

a, Raw ribo-seq reads for *unc-54(+)* (blue) and *unc-54(cc3389)* (green) animals, plotted as read pile-ups. Mismatched bases are indicated with black bars. Location of the normal stop codon and the first in-frame stop codon are indicated with 'TAA' and dotted lines. The extension in *unc-54(cc3389)* is 30 amino acids. **b**, The number of ribo-seq reads in the last 30 codons, compared to the previous 30 codons, for all mRNAs. Linear regression was performed on all points (solid line), and twofold difference shown (dashed lines). **c**, The distribution of ribo-seq reads in the last 30 codons (90 nt) of *unc-54(cc3389)* is shown in green, and the

95% confidence interval (CI) for all open reading frames in dashed lines. **d**, The fraction of in-frame ribo-seq reads in the last 30 codons is plotted as a function of read counts in the last 30 codons, and *unc-54(cc3389)* highlighted. **e**, The distribution of read lengths in the last 30 codons of *unc-54(cc3389)*, and all open reading frames (95% confidence interval, dashed lines). For **b–d**, reads were restricted to 28, 29, 30 nt lengths. For **b–e**, a 12 nt offset was performed for the ribosomal P-site, and read counts were derived solely from the *unc-54(cc3389)* ribo-seq library. For **c** and **e**, a minimum 15 read counts was imposed to obtain the 95% confidence interval from 'all genes'.

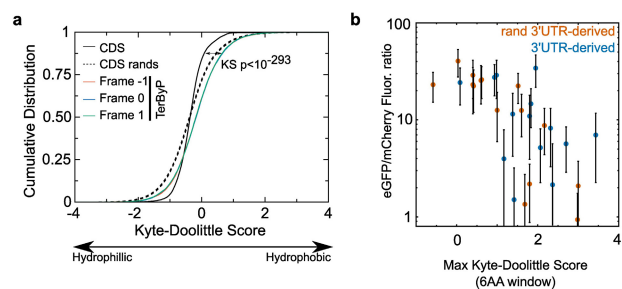


Extended Data Figure 7 | Lack of general conservation of coding potential downstream of stop codons in *Caenorhabditis*. Whole-genome alignment of six nematode species with *C. elegans* genome assembly ce10/WS220 was obtained from the UCSC genome browser. For each annotated transcript, the aligned bases from the multiple species alignment were extracted and compared to the reference (*C. elegans*) genome. The left plot shows summary information of the alignment centred on annotated stop codons; the right plot shows the same centred on the first in-frame stop codon in 3' UTRs. In red is the substitution frequency, that is, the number of mismatched bases divided by the number of aligned bases at a given position. The enrichment of 'wobble' position mutations is apparent as an increase in substitutions at the third position of each codon in the CDS. In green is the synonymous substitution frequency, that is, for codons beginning at a given position, the fraction of mutations that yield a synonymous substitution divided by all mutations at that position (synonymous and non-synonymous). The tendency to conserve amino acids in the CDS is apparent as a green spike at every in-frame codon. The change in substitution frequency and synonymous substitution frequency about the first in-frame stop codon (right plot) is due to a tendency for NTR codons to be conserved, and for AAN/AGN/GAN codons to not be conserved in 3' UTRs, regardless of frame.



Extended Data Figure 8 | Nucleotide and amino acid composition of readthrough regions (*C. elegans*). CDSs and 3' UTRs were analysed for various sequence properties. For simplicity, only genes and 3' UTRs for which a single 3' UTR was annotated were considered. Similar results were obtained with genes with multiple 3' UTRs. **a**, Nucleotide frequency of CDS, 3' UTR, and TerByP (region between annotated stop codon and first in-frame stop codon). **b**, Frequency of amino acids in all three possible frames for the TerByP region. 3' UTRs were translated one codon past the stop codon of the CDS until the next in-frame stop codon, with nonstop 3' UTRs ignored. Highlighted are codons with high G content (GGN, Gly) and high T content (TTY, Phe). **c**, TerByP regions tend to be hydrophobic, regardless of frame. Kyte–Doolittle score was used as a measure of hydrophobicity⁵⁴. To reduce noise, only TerByP regions at least 10 amino acids long were considered. *P* value is from Kolmogorov–Smirnov test comparing CDSs and TerByP sequences (each frame has *P* value < 10⁻²⁹³ for this comparison). As the TerByP sequences are shorter than CDSs on average, the distribution of TerByP hydrophobicity scores will tend to have higher variance than CDSs. Random portions of CDSs were taken, length-matched to TerByP frame zero peptide lengths. This was repeated

100 times, and the 95% confidence interval is shown (dashed lines, 'CDS rands'). **d**, Hydrophobicity of the inserts is correlated with a negative effect on GFP fluorescence. The GFP:mCherry fluorescence ratio (Fig. 2b) was plotted against the maximum Kyte–Doolittle score in a six amino acid window for each insert. (Similar results were obtained using the Kyte–Doolittle score averaged across the entire sequence.) Mean (circle) and s.d. (bars) are shown. 3'-UTR-derived sequences are in blue, and non-3'-UTR-derived sequences are in red. To avoid redundancy or skewing of the data, in cases where multiple constructs were present with the same peptide sequence (for example, *unc-54(TerByP)*, *unc-54(TerByP, syn1)*, and *unc-54(TerByP, syn2)*), only the first of these was used. **e**, Hydrophobicity analysis of the TerByP extensions obtained by CRISPR–Cas9 engineering at the *unc-22* and *unc-54* loci. '+1/-1 TerByP' denotes the gain or loss of a nucleotide, generating a late frameshift and allowing translation to proceed past the annotated stop codon out of frame with the upstream open reading frame. In each case, Kyte–Doolittle hydrophobicity was used to analyse the C-terminal appendage. The least phenotypically affected strain of the three is shown in bold.



Extended Data Figure 9 | Nucleotide and amino acid composition of readthrough regions (*H. sapiens*). **a, b**, Similar analysis of hydrophobicity as in Extended Data Fig. 8c, d, performed in humans.

Extended Data Table 1 | Translation into 3' UTRs at endogenous loci tends to yield hypomorphs

Gene	Genotype	Phenotype	Isolates
<i>pha-4</i>	<i>pha-4(TerByP)</i>	wild type	2
<i>unc-22</i>	<i>unc-22(TerByP)</i>	wild type	2
	<i>unc-22(-1, TerByP)</i>	twitcher	1
	<i>unc-22(+1, TerByP)</i>	twitcher	1
	<i>unc-22(unc-22::GFP)</i>	wild type	3
<i>unc-45</i>	<i>unc-45(TerByP)</i>	emb lethal	3
	<i>unc-45(3xFLAG::TEV::3xHA)</i>	wild type	2
<i>tra-2</i>	<i>tra-2(TerByP)</i>	XX males	2
	<i>tra-2(3xFLAG)</i>	wild type	2
	<i>tra-2(3xFLAG::TEV::3xHA)</i>	XX males	2
<i>unc-54</i>	<i>unc-54(TerByP)</i>	paralyzed	2
	<i>unc-54(unc-54::gfp)</i>	wild type	2
	<i>unc-54(+1, TerByP)</i>	paralyzed	1
	<i>unc-54(-1, TerByP)</i>	weak Unc	1

CRISPR–Cas9 editing¹⁷ was used to construct the mutations shown. See Supplementary Table 1 for precise nucleotide sequences of all strains. ‘–1/+1 TerByP’ indicate the loss or gain of one nucleotide relative to the zero frame, generating a frameshift over the stop codon, and translation into the 3' UTR out of frame with the coding sequence. For *unc-22*, ‘wild type’ indicates a lack of twitching, even in 1 mM levamisole. For *unc-54(-1, TerByP)*, ‘weak Unc’ animals were visibly slower than *unc-54(+)*, but faster than *unc-54(TerByP)*. All mutant phenotypes were recessive.

Cryo-EM structure of a human cytoplasmic actomyosin complex at near-atomic resolution

Julian von der Ecken¹, Sarah M. Heissler^{2†}, Salma Pathan-Chhatbar², Dietmar J. Manstein^{2,3} & Stefan Raunser¹

The interaction of myosin with actin filaments is the central feature of muscle contraction¹ and cargo movement along actin filaments of the cytoskeleton². The energy for these movements is generated during a complex mechanochemical reaction cycle^{3,4}. Crystal structures of myosin in different states have provided important structural insights into the myosin motor cycle when myosin is detached from F-actin^{5–7}. The difficulty of obtaining diffracting crystals, however, has prevented structure determination by crystallography of actomyosin complexes. Thus, although structural models exist of F-actin in complex with various myosins^{8–11}, a high-resolution structure of the F-actin–myosin complex is missing. Here, using electron cryomicroscopy, we present the structure of a human rigor actomyosin complex at an average resolution of 3.9 Å. The structure reveals details of the actomyosin interface, which is mainly stabilized by hydrophobic interactions. The negatively charged amino (N) terminus of actin interacts with a conserved basic motif in loop 2 of myosin, promoting cleft closure in myosin. Surprisingly, the overall structure of myosin is similar to rigor-like myosin structures in the absence of F-actin, indicating that F-actin binding induces only minimal conformational changes in myosin.

A comparison with pre-powerstroke and intermediate (P_i-release)⁷ states of myosin allows us to discuss the general mechanism of myosin binding to F-actin. Our results serve as a strong foundation for the molecular understanding of cytoskeletal diseases, such as autosomal dominant hearing loss and diseases affecting skeletal and cardiac muscles, in particular nemaline myopathy and hypertrophic cardiomyopathy.

Using electron cryomicroscopy (cryo-EM) and single-particle-based analysis of helical specimens (Methods), we determined the structure of a human actomyosin–tropomyosin (ATM) complex, composed of the motor domain of non-muscular myosin-2C (NM-2C), cytoplasmic γ 1-F-actin and cytoplasmic tropomyosin 3.1 (Fig. 1, Extended Data Figs 1a–g and 2 and Supplementary Video 1). We also reprocessed our previous F-actin–tropomyosin data set¹² and obtained an improved reconstruction at 3.6 Å resolution (Extended Data Figs 1h–k and 2a). The density of tropomyosin did not improve in both data sets and is limited to ~ 7 Å as described previously¹².

The ATM structure reveals that myosin interacts intimately with F-actin (Fig. 1a). The overall organization of the ATM complex is similar to that described in our previous structure of the complex

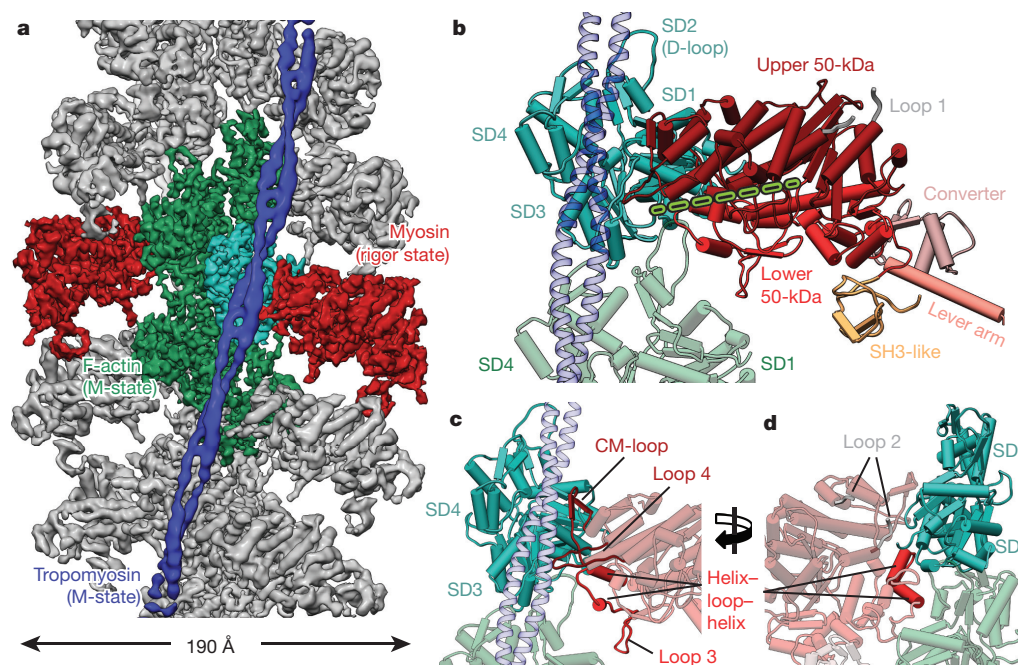


Figure 1 | Structure of the ATM complex. **a**, Cryo-EM reconstruction of F-actin (five central subunits in green and cyan) decorated with tropomyosin (blue) and myosin (central molecules in red). The peripheral densities (shown in grey) and tropomyosin were low-pass filtered and symmetrized for better visualization. **b**, Subdomain organization of F-actin and myosin head region. The closed actin-binding cleft between L50 (red) and U50 (dark red) domains is indicated with a dotted line. **c**, **d**, Front and back views of the F-actin–myosin interface. Involved structural parts of myosin are highlighted in red. For all figures and videos, we use a general colour code for each protein and state, if not labelled differently. The central F-actin subunit is shown in cyan (M-state) and yellow (A-state); surrounding F-actin subunits are depicted in green (M-state) and in darker yellow (A-state); rigor state myosin (red), P_i-release state myosin (blue) and PPS state myosin (purple). Less relevant parts of models or densities in respective figures are depicted in grey or faded out.

¹Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ²Institute for Biophysical Chemistry, Hannover Medical School, 30625 Hannover, Germany. ³Division for Structural Analysis, Hannover Medical School, 30625 Hannover, Germany. [†]Current address: Laboratory of Molecular Physiology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

between *Dictyostelium discoideum* myosin-IE, skeletal α -F-actin and α -tropomyosin¹⁰ (Supplementary Videos 2 and 3). However, given the superior resolution of our present structure, we could clearly identify large side chains and therefore reveal most intermolecular interactions between F-actin and myosin in detail.

The large interface between the two proteins is formed mainly between the helix–loop–helix (HLH) motif and surface loops of myosin (CM-loop, loop 2, loop 3, loop 4, and ‘activation’ loop¹³) and the subdomain (SD) 1 and 2 of one actin subunit and SD2 (D-loop) of the adjacent actin subunit as previously predicted^{9–11} (Fig. 1b–d and Supplementary Video 3).

Tpm3.1, which is resolved to ~ 7 Å, is in the same position as skeletal muscle tropomyosin 1.1 (Tpm1.1) in our previous structure, namely the M-state¹⁰, interacting with loop 4 of myosin and SD3 of actin (Figs 1b, c and 2a). Interestingly, although Tpm3.1 is shorter (stretching over six actin subunits) than Tpm1.1 (stretching over seven actin subunits), the pitch of the coiled-coil structure is equivalent. Tpm3.1, which is mainly negatively charged on its surface, interacts with arginine 384 of loop 4, indicating electrostatic interactions (Fig. 2a). In addition, the negatively charged residue D387 on loop 4 (N377–D393) interacts with a positively charged region on F-actin (K325, K327) (Fig. 2b). Notably, in the absence of myosin, these actin residues directly interact with tropomyosin (A-state)¹².

The HLH motif (L550–E575) in the lower 50-kDa (L50) domain of NM-2C plays an essential role in strong binding of myosin to F-actin (Figs 1c, d and 2c, d). It enters a hydrophobic groove on actin that is formed between two adjacent actin subunits comprising SD1 and SD3 of one and the D-loop (R38–V53) of the adjacent subunit. In particular, the hydrophobic loop of the HLH motif interacts with the hydrophobic groove and F560 is completely immersed into a hydrophobic cavity resembling a lock-and-key interaction (Fig. 2d, Extended Data Fig. 3a–d and Supplementary Video 4). The key role of F560 for the actomyosin interaction has also been shown by mutational analyses in which a F560A mutation resulted in a complete disruption of motility, whereas alanine mutants of the directly adjacent W559 and P561 showed only one-tenth the motility compared with wild type¹⁴. Interestingly, compared with its position in the pre-powerstroke (PPS) state the loop is the only part of the HLH motif that alters its position upon actin binding, stressing its important role in the actin–myosin interaction (Extended Data Fig. 3a, c).

In addition to the hydrophobic contacts, there are also two electrostatic interactions that stabilize the HLH motif binding to F-actin. E570 probably forms a salt bridge with K49 of the D-loop and E556 interacts with the backbone of S349 and T350 in SD1 of F-actin (Fig. 2c, Extended Data Fig. 3b, e and Supplementary Video 4). Both residues are part of a highly conserved acidic patch in several myosin classes (Extended Data Fig. 3e, f), and a single point mutation (E556Q) in myosin results in a tenfold reduced F-actin binding affinity¹⁵.

The cardiomyopathy loop (CM-loop), forming one antiparallel β -strand pair (T417–T432), is the major site of the myosin upper 50-kDa (U50) domain that interacts with actin (Figs 1c, and 2e, f and Supplementary Video 4). The CM-loop is fully ordered and the interface with actin is mainly stabilized by hydrophobic interactions (Fig. 2f), supported by weak electrostatic interactions at the tip and the base of the CM-loop (Fig. 2e). K429, which is found in all myosin-II isoforms (Extended Data Fig. 4a, b), interacts with a negatively charged patch on F-actin formed mainly by D24 and E333 in SD1 and SD3, respectively (Fig. 2e). In addition, the positively charged tip (R424) interacts with the negatively charged region around E55 and D92 of F-actin (Fig. 2e). However, residue 424 is only positively charged in smooth and non-muscular isoforms of myosin (Extended Data Fig. 4b) and therefore does not play a role in skeletal and cardiac muscles. Importantly, we did not find any prominent possible salt bridges that would stabilize the contacts between the CM-loop and F-actin, supporting previous mutagenesis studies suggesting that charged

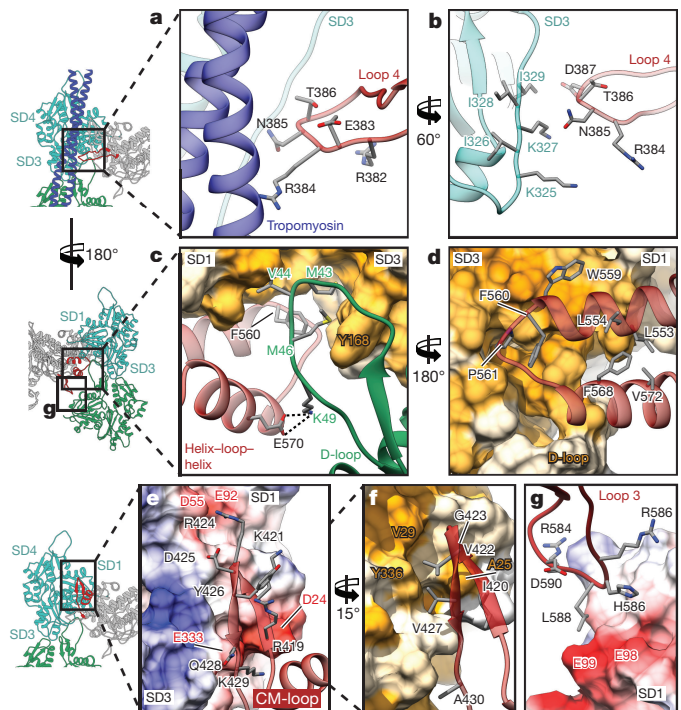


Figure 2 | Interfaces of the ATM complex. **a, b**, Interface of loop 4 with tropomyosin (blue) and SD3 of one F-actin subunit (cyan). **c, d**, Interaction of the HLH motif of myosin (red) with hydrophobic groove formed by the D-loop of one F-actin subunit (green ribbon in **c** and as surface in **d**) and SD1 and SD3 of the adjacent subunit (surfaces depicted by low (white) to high (yellow) hydrophobicity). Hydrophobic residues at the interface (**c, d**) and a possible electrostatic interaction (dotted lines (**c**)) are highlighted. **e, f**, The CM-loop (red) binds to a region formed by SD1 and SD3 of F-actin. Charged (**e**) and hydrophobic (**f**) residues of myosin. **g**, Loop 3 (red) interacts with SD1 of an adjacent F-actin subunit forming the Milligan contact^{9,21}. The F-actin surface is coloured either by hydrophobicity (**f**) or electrostatic Coulomb potential from -10 kcal mol⁻¹ (red) to $+10$ kcal mol⁻¹ (blue) (**e, g**). In all panels, coloured residue labels depict F-actin residues.

residues, in particular the highly conserved residue D425, play a minor role in this interface^{14,16}.

As speculated previously⁸, the highly conserved and disease-related residue R419 (R403 in β -cardiac myosin) indeed does not directly interact with F-actin (Fig. 2e, Extended Data Fig. 4b–d and Supplementary Video 4). In our structure, R419 clearly interacts with Y426 on the opposing strand of the CM-loop, thereby bridging and stabilizing the conformation of the loop (Fig. 2e). Both residues are highly conserved, suggesting that this bridge is present in all myosin-II isoforms (Extended Data Fig. 4b). Many disease-causing mutations are found in the region of the CM-loop, demonstrating the high importance of the CM-loop for the strong binding between actin and myosin (Extended Data Fig. 4b–e).

Several studies showed that loop 2, connecting the L50 and U50 domains in myosin (W638–T669), plays a major role in the initial binding to F-actin^{14,15,17–19}. We see clear density for loop 2 in the ATM structure. However, whereas the base of the loop is ordered, the rest of the loop is more flexible (Fig. 3a, b). It occupies a large predominantly hydrophobic surface of the actin SD1 domain (Fig. 3c and Extended Data Fig. 5a, b).

A conserved hydrophobic patch at the carboxy (C)-terminal base of loop 2 (G665–F667) and the tip of helix-R (W559) (Fig. 3c, Extended Data Fig. 3f and Supplementary Video 4) interacts with a hydrophobic groove of actin SD1 (Fig. 3c and Extended Data Fig. 3d). Mutagenesis studies showed that especially W559 is essential for forming the F-actin–myosin interface and obtaining motility^{14,20}. Compared with

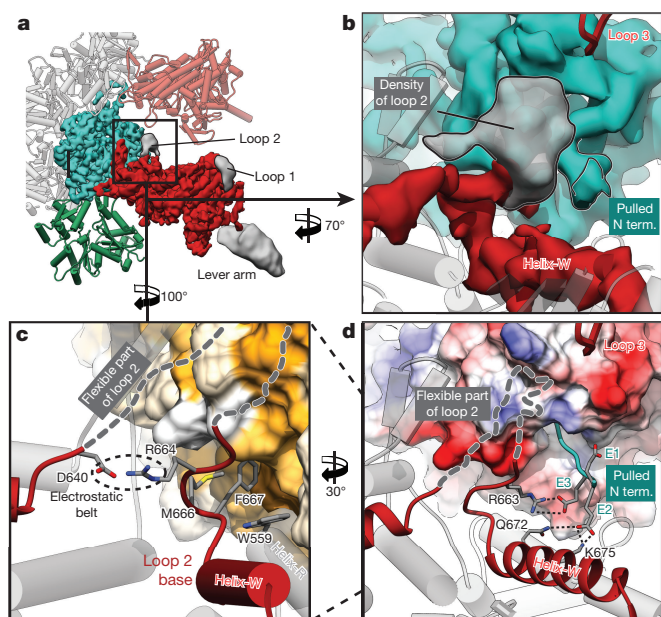


Figure 3 | Stabilization of loop 2. **a**, Myosin densities of lower resolution (grey) are shown together with more highly resolved regions of myosin (red) and F-actin (cyan). The densities can be assigned to more flexible parts of loop 1, loop 2 and the outer lever arm (Methods), respectively. **b**, Close-up view onto region of the density map corresponding to loop 2. Loop 3 of the adjacent myosin does not interact with loop 2 (red). **c**, **d**, Interaction of the stabilized base of loop 2 with SD1 of F-actin (coloured by hydrophobicity (**c**) or by electrostatic Coulomb potential (**d**)). The flexible part of loop 2 and possible electrostatic interactions are indicated by different dotted lines.

their position in the PPS state, these residues orient towards actin to stabilize the newly formed interface (Fig. 4a).

Interestingly, the adjacent conserved positively charged region (R661–R664) interacts with the acidic N terminus of actin and a negatively charged area (D23, D24) of SD1 (Fig. 3d, Extended Data Fig. 5b and Supplementary Video 4). While one of the conserved arginines (R663) forms a possible salt bridge with actin E3 (Fig. 3d), the other one (R664) interacts with D640 and D642 at the other end of loop 2, forming an electrostatic belt that stabilizes the base of loop 2 (Fig. 3c). Notably, D640 and D642 are only conserved in myosins with a long loop 2 (Extended Data Fig. 5c), suggesting that the electrostatic belt is not required for myosins with a shorter loop 2.

In line with our previous observation¹⁰, loop 3 (Q576–D593) forms the so-called Milligan-contact^{9,21} connecting the L50 domain to SD1 of the adjacent actin subunit (Figs 1c and 2g and Extended Data Fig. 5d, e). The small interface is only formed by complementary charged surfaces and not by specific salt bridges as previously expected (Fig. 2g). The weak nature of the interactions, and the fact that not all myosin isoforms have a long loop 3 that can form this contact²², suggest that, for most myosin proteins, loop 3 plays only an ancillary role in strong F-actin binding.

Several studies suggested that a proline-rich loop in the L50 domain, a so-called activation loop (I541–G549), is directly involved in activation of myosin by interacting with the negatively charged N terminus of actin^{13,23}. Our ATM structure confirms that this loop is part of the actomyosin interface. Together with helix-W and the base of loop 2, it forms a positively charged basin that interacts with the negatively charged N terminus of actin (Fig. 3d, Extended Data Fig. 6a, b and Supplementary Video 4). However, R543, which is the only positively charged residue of the loop and conserved in myosin-II (Extended Data Fig. 6c), points away from the interface and therefore cannot be involved in a direct interaction with the glutamates of the actin N terminus (Fig. 4b and Extended Data Fig. 6d). In other myosins with shorter

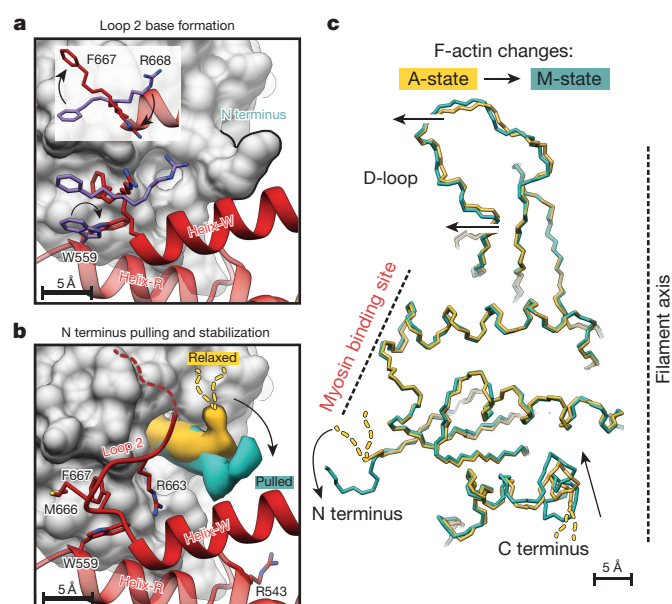


Figure 4 | Comparison of PPS and rigor state and induced changes in F-actin. **a**, Involved residues of loop 2 and helix-R undergo changes in rotamer orientation during base stabilization of loop 2 (PPS state (purple), PDB accession number 5I4E; rigor state (red)). The F-actin surface is shown in grey. Inset shows a top view of rotating residues of loop 2. F667 rotates to SD1, R668 rotates outwards to the U50 of myosin. **b**, Density comparison of the N terminus between myosin-unbound F-actin (yellow, A-state) and myosin-bound F-actin (cyan, M-state) illustrates the pulled conformation of the terminus induced by loop 2 interaction. The flexible N terminus (1–4 in α -actin) in the A-state is depicted as dotted lines. **c**, Superposition of SD1 and SD2 of F-actin in the A-state (yellow) and M-state (cyan) visualizing the myosin-induced changes in F-actin (indicated by arrows).

activation loops, namely myosin-V, myosin-VI or *Dictyostelium* myosin-II, the position of the positively charged residue is shifted by one position and would therefore allow a direct interaction with the N terminus (Extended Data Fig. 6c, e, f). Compared with the PPS conformation, the proline-rich loop orients slightly closer to F-actin (Extended Data Fig. 6b). Since the actin-induced conformational changes in the proline-rich loop and the adjacent relay helix are minor, we conclude that they are probably not responsible for a direct activation of myosin and therefore suggest using the term supporting loop rather than activation loop.

To identify myosin-induced conformational changes in F-actin, we compared the ATM structure with our reprocessed F-actin–tropomyosin structure. As expected from our previous observations¹⁰, the overall changes are minimal. Whereas the D-loop and other interface regions orient slightly towards myosin (Fig. 4c and Extended Data Fig. 3a), areas interacting with the CM-loop and more distal regions of actin only move marginally away from the interface (Extended Data Fig. 7a–c).

The most prominent changes occur at the N and C termini of actin (Fig. 4b, c and Extended Data Fig. 7d–k). The highly conserved and negatively charged N terminus (Extended Data Fig. 7c), which is only partly resolved in F-actin¹², but essential for myosin binding^{24–26}, is completely ordered in the ATM structure and pulled into a positively charged basin on the myosin structure (Fig. 4b, c and Extended Data Fig. 6a). Two glutamates at its tip (E2, E3) are potentially involved in salt bridges with positively charged residues on helix-W and loop 2 (Fig. 3d). The conformational change of the N terminus is partly transmitted to the nucleotide-binding pocket (Extended Data Fig. 7e, f). Actin residues D10 and K17 slightly change their position (Extended Data Fig. 7f, g); however, this does not considerably alter the position of ADP and Mg^{2+} (Extended Data Fig. 7f, h).

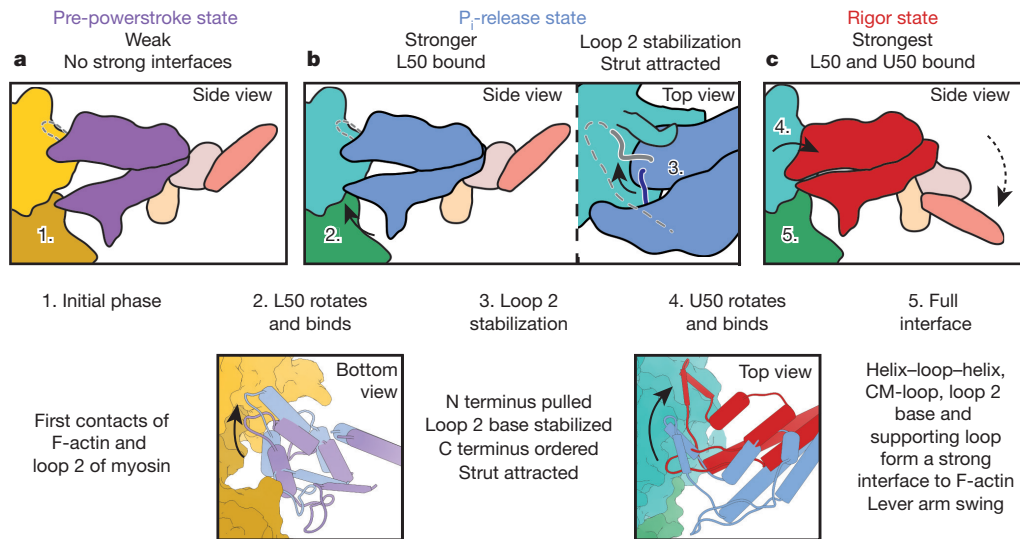


Figure 5 | Model of myosin binding to F-actin. a–c, Cartoon representation of the myosin–F-actin-binding mechanism. Numbers refer to different steps of the actomyosin interaction and are directly described in the lower panel. U50 and L50 domains are coloured according to their respective myosin state: PPS (purple, PDB accession number 5I4E), P_i -release (blue, homology model of PDB accession number 4PFO (ref. 7))

Interestingly, although the actin C terminus does not participate in forming the actomyosin interface, it is completely ordered in the ATM structure and orients towards the SD1 domain (Fig. 4c and Extended Data Fig. 7i–k). This is in line with previous studies showing that myosin binding to F-actin results in quenching of fluorescence of pyrene-labelled actin at C373 (C374 in α -actin)¹⁴. The higher myosin-induced stability of the D-loop, N terminus and surrounding regions is probably responsible for the stabilization of the C terminus, which is not well ordered in the F-actin–tropomyosin structure¹² (Extended Data Fig. 7j). We believe that the minimal but substantial myosin-induced conformational changes are exemplary for most actin–myosin interactions. We therefore suggest using the term M-state for F-actin bound to myosin in contrast to A-state for bare F-actin (Figs 1a and 4c).

To understand which conformational changes F-actin binding induces in myosin, we compared our rigor ATM structure with rigor-like crystal structures (Extended Data Fig. 8). Surprisingly, the overall structures of the different rigor-like states are very similar compared with that of the rigor state (Extended Data Fig. 8a, b), indicating that F-actin induces only minimal conformational changes in myosin as previously predicted^{11,27}. Differences are found at actin-interacting loops that are partly ordered in the crystal structures and are stabilized by actin in the ATM structure (Extended Data Fig. 8c). In addition, we found that the converter and lever arm regions differ in their position relative to the rest of the protein (Extended Data Fig. 8d). Importantly, the comparison between rigor and rigor-like structures shows that F-actin stabilizes the closed conformation of myosin, but does not induce major additional conformational changes.

To gain further insight into how actin accommodates first a weak and then a strong myosin-binding state, we compared our rigor NM-2C structure with the crystal structure of the same protein in the PPS state. On the basis of recently determined crystal structures of the motor domain of myosin-VI in the P_i -release state, an important intermediate state, a detailed mechanism of myosin binding to F-actin has been proposed⁷. To obtain the best possible approximation for an F-actin–NM-2C complex in the P_i -release state, we used a two-step approach. First, we calculated a homology model of the P_i -release state of NM-2C on the basis of the atomic model of myosin-VI (ref. 7). On the basis of the positions of actin and myosin in our rigor actomyosin structure and the myosin PPS structure, we then performed different alignments

and rigor state (red). Other domains are coloured as in Fig. 1b. F-actin subunit colours give their current state (A-state, yellow; M-state, cyan and green). Loop 2 (grey) and strut (dark blue) are shown as lines. Flexibility in loop 2 is indicated by dotted lines. Arrows highlight rotations and binding of domains.

of the P_i -release state model to obtain a model for the actomyosin complex in its P_i -release state (Fig. 5 and Extended Data Fig. 9a–c). Exclusive alignment to the U50 domain in all states would cause steric clashes with F-actin in the P_i -release state (Extended Data Fig. 9a). Alignment to the L50 domain would result in a rotation of the U50 away from F-actin before the final strong binding (Extended Data Fig. 9b). We therefore chose a combined alignment instead (Extended Data Fig. 9c and Supplementary Video 5) to describe the possible global conformational changes during myosin binding (Extended Data Fig. 9d–f, Fig. 5 and Supplementary Video 6). We aligned the P_i -release state model first to the L50 domain of the rigor state. The PPS state was then aligned to the U50 domain of the P_i -release state.

There is general agreement that loop 2 initiates weak binding of myosin–ADP– P_i to F-actin by interacting with the SD1 and SD3 of one actin subunit^{14,15,17–19} (Extended Data Fig. 9d). This brings both the L50 and U50 close to F-actin (Fig. 5a). As proposed in ref. 7, the L50 domain rotates and binds to actin, resulting in the P_i -release state that represents the initial strong binding state of myosin. The interface is mainly mediated by hydrophobic interactions of the HLH domain with SD1, SD3 and the D-loop and hydrophilic interactions with the N terminus of actin, in line with our previous prediction¹⁰ (Fig. 5b). On the basis of our results, we propose that this process stabilizes the base of loop 2 (Extended Data Fig. 9e, Fig. 5b and Supplementary Video 6), creating a positively charged patch to which the negatively charged strut (bridge between L50 and U50) is attracted (Extended Data Fig. 9f and Fig. 5b). Thus, the base of loop 2 acts as a key region that shifts the equilibrium between the open and closed actin-binding cleft²⁸ towards the closed conformation, in which F-actin directly interacts with the CM-loop and loop 4 of the U50 domain (Fig. 5c). Notably, both the strut and the highly conserved basic region at the base of loop 2 have previously been shown to be essential for strong binding^{17,27,29}. In our model, the closure of the cleft is mediated by a rotation of the U50 domain towards F-actin and not by a back-rotation of the L50 domain (Fig. 5c).

As described before (reviewed in ref. 4), cleft closure results in the strong binding of myosin to actin, providing the necessary anchoring of the motor domain for the subsequent powerstroke (Fig. 5c). The time point and effect of P_i and ADP release during this process is highly debated^{7,30}. Because we lack high-resolution structures of intermediate states of myosin bound to F-actin, we cannot determine whether P_i is

released before or after the powerstroke. However, our observation that the rigor state is very similar to the rigor-like state suggests that actin promotes and stabilizes the closed conformation of myosin, ultimately resulting in the release of phosphate and ADP.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 February; accepted 29 April 2016.

Published online 20 June 2016.

1. Huxley, H. E. The mechanism of muscular contraction. *Science* **164**, 1356–1366 (1969).
2. Heissler, S. M. & Manstein, D. J. Nonmuscle myosin-2: mix and match. *Cell. Mol. Life Sci.* **70**, 1–21 (2013).
3. Lynn, R. W. & Taylor, E. W. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* **10**, 4617–4624 (1971).
4. Sweeney, H. L. & Houdusse, A. Structural and functional insights into the myosin motor mechanism. *Annu. Rev. Biophys.* **39**, 539–557 (2010).
5. Coureux, P.-D., Sweeney, H. L. & Houdusse, A. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO J.* **23**, 4527–4537 (2004).
6. Gulick, A. M., Bauer, C. B., Thoden, J. B. & Rayment, I. X-ray structures of the MgADP, MgATP γ S, and MgAMPPNP complexes of the *Dictyostelium discoideum* myosin motor domain. *Biochemistry* **36**, 11619–11628 (1997).
7. Llinas, P. *et al.* How actin initiates the motor activity of myosin. *Dev. Cell* **33**, 401–412 (2015).
8. Lorenz, M. & Holmes, K. C. The actin-myosin interface. *Proc. Natl Acad. Sci. USA* **107**, 12529–12534 (2010).
9. Rayment, I. *et al.* Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–65 (1993).
10. Behrmann, E. *et al.* Structure of the rigor actin-tropomyosin-myosin complex. *Cell* **150**, 327–338 (2012).
11. Holmes, K. C., Angert, I., Kull, F. J., Jahn, W. & Schröder, R. R. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* **425**, 423–427 (2003).
12. von der Ecken, J. *et al.* Structure of the F-actin-tropomyosin complex. *Nature* **519**, 114–117 (2015).
13. Várkuti, B. H. *et al.* A novel actin binding site of myosin required for effective muscle contraction. *Nature Struct. Mol. Biol.* **19**, 299–306 (2012).
14. Onishi, H., Mikhailenko, S. V. & Morales, M. F. Toward understanding actin activation of myosin ATPase: the role of myosin surface loops. *Proc. Natl Acad. Sci. USA* **103**, 6136–6141 (2006).
15. Furch, M., Rammel, B., Geeves, M. A. & Manstein, D. J. Stabilization of the actomyosin complex by negative charges on myosin. *Biochemistry* **39**, 11602–11608 (2000).
16. Sasaki, N., Asukagawa, H., Yasuda, R., Hiratsuka, T. & Sutoh, K. Deletion of the myopathy loop of *Dictyostelium* myosin II and its impact on motor functions. *J. Biol. Chem.* **274**, 37840–37844 (1999).
17. Joel, P. B., Trybus, K. M. & Sweeney, H. L. Two conserved lysines at the 50/20-kDa junction of myosin are necessary for triggering actin activation. *J. Biol. Chem.* **276**, 2998–3003 (2001).
18. Murphy, C. T. & Spudis, J. A. The sequence of the myosin 50–20K loop affects myosin's affinity for actin throughout the actin-myosin ATPase cycle and its maximum ATPase activity. *Biochemistry* **38**, 3785–3792 (1999).
19. Uyeda, T. Q. P., Ruppel, K. M. & Spudis, J. A. Enzymatic activities correlate with chimaeric substitutions at the actin-binding face of myosin. *Nature* **368**, 567–569 (1994).
20. Kojima, S. *et al.* Functional roles of ionic and hydrophobic surface loops in smooth muscle myosin: their interactions with actin. *Biochemistry* **40**, 657–664 (2001).
21. Milligan, R. A., Whittaker, M. & Safer, D. Molecular structure of F-actin and location of surface binding sites. *Nature* **348**, 217–221 (1990).
22. Van Dijk, J., Furch, M., Lafont, C., Manstein, D. J. & Chaussepied, P. Functional characterization of the secondary actin binding site of myosin II. *Biochemistry* **38**, 15078–15085 (1999).
23. Várkuti, B. H., Yang, Z. & Malnasi-Csizmadia, A. Structural model of weak binding actomyosin in the prepowerstroke state. *J. Biol. Chem.* **290**, 1679–1688 (2015).
24. Cook, R. K., Root, D., Miller, C., Reisler, E. & Rubenstein, P. A. Enhanced stimulation of myosin subfragment 1 ATPase activity by addition of negatively charged residues to the yeast actin NH2 terminus. *J. Biol. Chem.* **268**, 2410–2415 (1993).
25. Miller, C. J., Wong, W. W., Bobkova, E., Rubenstein, P. A. & Reisler, E. Mutational analysis of the role of the N terminus of actin in actomyosin interactions. Comparison with other mutant actins and implications for the cross-bridge cycle. *Biochemistry* **35**, 16557–16565 (1996).
26. Sutoh, K., Ando, M., Sutoh, K. & Toyoshima, Y. Y. Site-directed mutations of *Dictyostelium* actin: disruption of a negative charge cluster at the N terminus. *Proc. Natl Acad. Sci. USA* **88**, 7711–7714 (1991).
27. Coureux, P.-D. *et al.* A structural state of the myosin V motor without bound nucleotide. *Nature* **425**, 419–423 (2003).
28. Klein, J. C. *et al.* Actin-binding cleft closure in myosin II probed by site-directed spin labeling and pulsed EPR. *Proc. Natl Acad. Sci. USA* **105**, 12867–12872 (2008).
29. Sasaki, N., Ohkura, R. & Sutoh, K. Insertion or deletion of a single residue in the strut sequence of *Dictyostelium* myosin II abolishes strong binding to actin. *J. Biol. Chem.* **275**, 38705–38709 (2000).
30. Muretta, J. M., Rohde, J. A., Johnsrud, D. O., Cornea, S. & Thomas, D. D. Direct real-time detection of the structural and biochemical events in the myosin power stroke. *Proc. Natl Acad. Sci. USA* **112**, 14272–14277 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Hofnagel for assistance in cryo sample preparation. We acknowledge R. Matadeen and S. de Carlo for image acquisition at the Netherlands Centre for Nanoscopy in Leiden. We thank R. S. Goody for reading the manuscript. This work was supported by the Max Planck Society, the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) (grant number 615984) (to S.R.), the Behrens-Weise foundation (to S.R.) and German Research Foundation (DFG) grant MA 1081/21-1 (to D.J.M.). J.v.d.E. is a fellow of Studienstiftung des deutschen Volkes.

Author Contributions D.J.M. and S.R. designed the project. S.M.H. and S.P.-C. purified actin, tropomyosin, and myosin constructs. D.J.M. supervised protein work. J.v.d.E. prepared specimens, recorded, analysed and processed the data, and prepared figures. S.R. managed the project. J.v.d.E. and S.R. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The coordinates and electron microscopy density maps have been deposited in the Protein Data Bank (PDB) under accession numbers 5JLF and 5JLH and the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-8162 to EMD-8165. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stefan.raunser@mpi-dortmund.mpg.de).

Reviewer Information *Nature* thanks E. Nogales, J. Löwe and A. Houdusse for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification. G-actin (γ 1-actin, *ACTG1* from *Homo sapiens*) was recombinantly expressed using the baculovirus/Sf9-cell system and purified as described previously^{31,32}. Afterwards, the sample was polymerized to F-actin by increasing the salt concentration to 100 mM KCl and 2 mM MgCl₂. Tropomyosin 3.1 (*TPM3* from *H. sapiens*, isoform Tpm3.1) was expressed and purified from *Escherichia coli* on the basis of the protocol of ref. 33 with no additional modifications. The motor domain of non-muscular myosin-2C (*MYH14*, isoform 2 from *H. sapiens*) consisting of amino acids 1–799 was directly fused to an artificial lever arm (spectrin repeats 1 and 2 from α -actinin)³⁴ and a C-terminal Flag-tag. The protein was recombinantly overproduced in the Sf9/baculovirus system as previously described³⁴ and purified via Flag capture and size-exclusion chromatography on a Superdex 26/60–200 prep grade column. Before grid preparation for electron microscopy studies, the F-actin sample was spun down (100,000g) and carefully suspended in nucleotide-free F-actin buffer (5 mM Tris-HCl pH 7.5, 1 mM DTT, 100 mM KCl, and 2 mM MgCl₂). The actomyosin complex was prepared by mixing F-actin with tropomyosin initially at a molar ratio of 7:1. The final concentration of tropomyosin for frozen specimens was then adjusted empirically to obtain complete decoration and only little unbound tropomyosin in the background by standard negatively stained studies as described previously¹². The F-actin–tropomyosin filaments were decorated with myosin during preparation of vitrified sample grids (see below).

Optimized grid preparation and image acquisition for cryo-EM. Best conditions for reconstituting the F-actin–tropomyosin complex were screened by using the negatively staining protocol as described above. When myosin was incubated in solution with the optimized F-actin–tropomyosin sample before applying the sample to the grid, we always obtained only bundles of fully myosin-decorated actin filaments. Because of this, we optimized the protocol to reconstitute the full actomyosin filaments with reduced bundling on a grid. Therefore, the normal cryo-preparation protocol was changed. First we applied 2 μ l of F-actin–tropomyosin solution to a glow-discharged holey carbon grid (C-flats 2/1, Protochips), incubated for 20 s and manually blotted from the backside for less than a second with filter paper. A thin layer of solution stayed on the grid and the filaments where pre-straightened in the holes. Afterwards, 1.5 μ l of myosin solution (3 μ M in F-actin buffer without nucleotide) were added directly on the grid, incubated for 10 s and then manually blotted for 5 s from the backside with filter paper (Whatman no. 5), before vitrification by plunging the grid into liquid ethane using a Cp3 plunger (Gatan).

Screening for the best sample and blotting conditions was performed on a JEOL JEM 3200FSC electron microscope equipped with a field emission gun and operated at a voltage of 200 kV. The omega in-column energy filter of the microscope was used to estimate best ice conditions (~70–100 nm thickness). Finally, a data set was taken with a spherical aberration-corrected FEI Titan Krios transmission electron microscope equipped with an extra-high brightness field emission gun (X-FEG) and operated at a voltage of 300 kV. Although the sample preparation protocol was optimized, we had to screen and choose usable grid squares extensively. Images were recorded with a back-thinned 4k \times 4k FEI Falcon 2 direct detection camera under minimal dose conditions using the automatic data collection software EPU (FEI). Within each selected grid hole, three different positions were imaged, each with a total exposure of 1 s and a frame recording time of 55 ms. Seven frames from 85 to 475 ms with a total dose of ~16 electrons per square ångström and one total average (integrated image) with an electron dose of ~35 electrons per square ångström were used for image processing. The used magnification of 125,000 (nominal magnification of 59,000) corresponds to a pixel size of 1.1 Å. The defocus range of the data set was 0.7–2.8 μ m (Extended Data Table 1).

Image processing of the cryo-EM data set. In total, we collected ~6,300 images in two sessions. Despite the extensive pre-screening of grid squares before starting automatic data collection (see above), we deleted ~68% of the recorded images because of bundled filaments, contaminations or bad ice quality. Resulting frames were aligned and afterwards summed up using motion correction³⁵. The drift-corrected averages were used for determination of defocus and astigmatism values with CTFFIND3³⁶. Filaments were manually selected (Extended Data Fig. 1a) and exported from the 1 s integrated images using sxhelixboxer in SPARX³⁷ without changing the orientation of the filaments to the y axis. A total of ~138,000 segments were extracted with a box size of 256 pixels and a boxing distance of 29 pixels (overlap ~90%). Thus, the approximate distance between them (~32 Å) slightly exceeded the rise of the helical assembly of actin (~27–28 Å). The same procedure was applied to the drift-corrected average and all individual frames. Afterwards all segments were transformed to RELION³⁸-readable image stack formats and initial metadata files were created for further refinement steps in RELION.

First, two-dimensional reference-free classification and sorting of bad classes from the integrated images led to a resulting data set of 126,000 particles (Extended Data Fig. 1b). The resulting data set was used in several rounds of three-dimensional auto-refinements with particles from the integrated images and local three-dimensional auto-refinements with the particles from the drift-corrected averages. The refinement showed an expected Gaussian distribution of projection direction around the filament axis (Extended Data Fig. 1d, e). To improve processing time, we limited the tilt angle afterwards (Extended Data Fig. 1d). Finally, we applied a particle-based movie refinement and frame weighting³⁹ and continued three-dimensional auto-refinements with the resulting contrast-enhanced particles. We did not make use of helical symmetry during refinement but masked F-actin and myosin in the outer regions to focus the refinement on the central parts. Initially, we applied a standard spherical mask (diameter 270 Å, Extended Data Fig. 1d) to the reconstruction. After a global refinement with the spherical mask, we continued with local refinement and a mask at the size of seven F-actin subunits and six myosin molecules. We used a preliminary model and the 'Colour zone' and 'Split Map' options in CHIMERA^{40,41} to extract the central part of the map from the current density map. From this part of the map we calculated a smoothed mask with the 'relion_mask_create' function in RELION. To evaluate the results, we again performed a two-dimensional classification with projection parameters derived from three-dimensional refinement (Extended Data Fig. 1c). In single class averages, we could already detect secondary structure elements (Extended Data Fig. 1f, g). As a final sorting step, we deleted particles, which were outliers with respect to their neighbouring segments coming from the same filament. Therefore, we kept track throughout the whole processing steps to know to which filament each particle belongs. This resulted in 118,000 particles, which we used for a local three-dimensional auto-refinement. In the final iteration, we applied a mask of the central five F-actin subunits and two central myosin molecules (coloured region in Fig. 1a). This mask was created as described above for the mask of seven F-actin subunits and six myosin molecules. Outer regions of the lever arm were also excluded.

Fourier shell correlation (FSC) analysis was performed within the central area (five F-actin subunits and two myosin molecules) of the volume, resulting in an average resolution of 3.9 Å (FSC_{0.143} criterion⁴²) for the F-actin–myosin electron density map (Extended Data Fig. 2a). The density map of F-actin–myosin was then sharpened using a negative *b* factor of ~200 Å² and filtered to its nominal resolution. Because tropomyosin was masked out during refinements as we did before¹², we filtered the tropomyosin density map to ~7 Å and merged it with the final F-actin–myosin map to obtain a map of the entire F-actin–myosin–tropomyosin complex.

Local resolution was estimated using ResMap⁴³ on the full density map without masking (Extended Data Fig. 2b), revealing a resolution gradient from higher (~3.5 Å, F-actin core) to lower (4.5 Å–5 Å, outer myosin domains) values, which could be a result of induced forces from the protruded lever arm (Extended Data Fig. 2c–e). We converted the local resolution map to absolute frequencies and applied a local filtering algorithm on the final map with sxfilterlocal in SPARX. To estimate and confirm the helical symmetry of the actomyosin complex, we used the symmetry search function of sxhelicon_utils in SPARX.

We reprocessed our previous F-actin–tropomyosin¹² data set with the same protocol as described above (Extended Data Fig. 1h–k) and obtained an improved reconstruction at an average resolution of 3.6 Å for the central five F-actin subunits (Extended Data Fig. 2a). During the helical reconstruction approach, several asymmetrical units are averaged by symmetrization. This results in a decrease of resolution in flexible and bent regions. However, using the single particle approach, we refine only on the central region, thereby probably decreasing the influence of flexibility induced by the bending of the filament. This probably improves the resolution of the resulting reconstruction.

For each data set, we provide two density maps in the Electron Microscopy Data Bank (EMDB) databank. One entry contains the map after 'post-processing' in RELION without masking, the other one the same map locally filtered, masked and merged with the filtered tropomyosin density map at its respective position (as described above).

Atomic model building and refinement. The central F-actin subunit (chain A) and the central myosin molecules (chain F, G) show all available contacts to adjacent chains (B, C, D, E) and were therefore used for further structural analysis. We used our previous F-actin model¹² as a starting model for F-actin and the highly homologous model of rigor-like NM-2B (PDB accession number 4PD3⁴⁴) for myosin. We performed homology modelling with the respective sequences of both proteins using MODELLER⁴⁵. Derived models were rigid-body fitted into the density map, using 'Fit in Map' and the map was segmented with 'Split Map' in CHIMERA^{40,41}. Additionally, models were flexibly fitted with iMODFIT⁴⁶ into the respective density maps. Finally, the flexibly fitted models were used to create a starting F-actin–myosin model. The electron density was converted to

structure factors with the CCP4 program suite⁴⁷. The model and the density map were then used for real space refinement and model building in COOT⁴⁸. Flexible parts of the N terminus (residues 1–49), loop 1 (residues 222–232), loop 2 (residues 640–661) and lever arm (residues 817–1039) of myosin were deleted from the model. Some parts (especially surface loops and termini) of F-actin and myosin were built *de novo*. Domains showing resolution lower than 4.5 Å (outer myosin regions) were not manually changed and only the backbone trace was optimized.

The resulting model was refined in REFMAC, using the modified version of the program for cryo-EM maps⁴⁹ and applying secondary structure and reference restraints derived by ProSMART⁵⁰. In addition, we added non-crystallographic symmetry restraints for F-actin (chain A–E) and myosin (chain F and G). To prevent overfitting, we first determined refinement settings by refining the model (atoms randomly displaced by 0.5 Å) only versus a density map belonging to one half of the data set and compared the FSC curves of the refined model with both half maps (Extended Data Fig. 2f). Finally, the model was refined versus the full map with the derived refinement parameters, which did not show significant differences in the two FSC curves (Extended Data Fig. 2f). We used MOLPROBITY⁵¹ to evaluate the resulting atomic model. The data statistics are given in Extended Data Table 1.

For the reprocessed F-actin–tropomyosin data set, we used the same refinement strategies as for the actomyosin data set. The final data statistics are given in Extended Data Table 1.

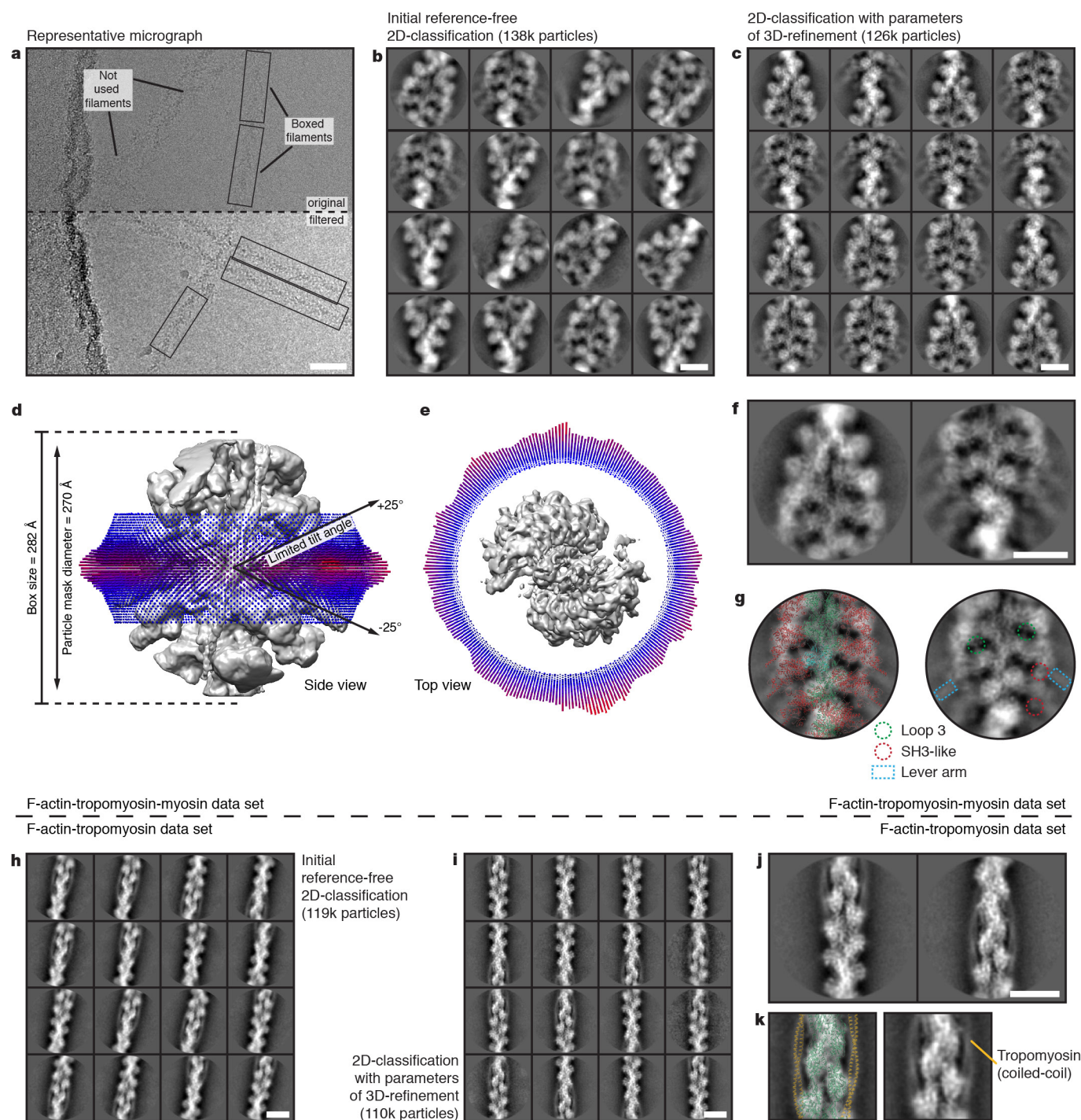
Fit of the tropomyosin models. To describe the interaction between F-actin–myosin and tropomyosin, we used the tropomyosin model from our previous structure¹⁰. As already described in the main text, the pitch of the coiled-coil structure is equivalent. We directly rigid-body fitted the tropomyosin model (PDB accession number 4A7F) into the density map, using ‘Fit in Map’ in CHIMERA⁴⁰. We reduced the model to the five central pseudo-repeats as before¹² and shifted the residue numbering regarding the differences from long Tpm1.1 to the shorter Tpm3.1. Owing to the limited resolution of the cryo-EM density in the region of tropomyosin, we avoided interpretation of tropomyosin at the single amino-acid level.

For the reprocessed F-actin–tropomyosin model, we could use the tropomyosin model from our previous model (PDB accession number 3JA8), as the structure and resolution in that region did not differ.

Structure analysis and visualization. For visualization of models and density maps in all figures and videos, we used CHIMERA⁴⁰. The actomyosin complex was protonated using H⁺⁺ (ref. 52) at pH 7.5, and the electrostatic Coulomb potential of the filament surface was calculated ranging from –10 to +10 kcal mol^{–1}. For visualization of the hydrophobicity per amino-acid residue, we used ‘Define attribute’ in CHIMERA and generated amino-acid-specific scores⁵³. The densities of flexible parts of myosin were detected by using the ‘Colour Zone’ function on a low-pass-filtered density map in CHIMERA⁴¹. For comparison of differences in density of A-state and M-state F-actin, we low-pass filtered both density maps to the same resolution of 3.9 Å. Sequence alignments were performed using the ClustalOmega online server⁵⁴. The HGMD⁵⁵ library and UNIPROT⁵⁶ were browsed to find mutations in regions of interest. For creating a homology model of NM-2C in the P_i-release state, we used MODELLER in the CHIMERA ‘Multalign viewer’-interface with NM-2C as target sequence and the coordinates of the previously determined crystal structure of the P_i-release state of myosin-VI (PDB accession number 4PFO)⁷ as reference structure.

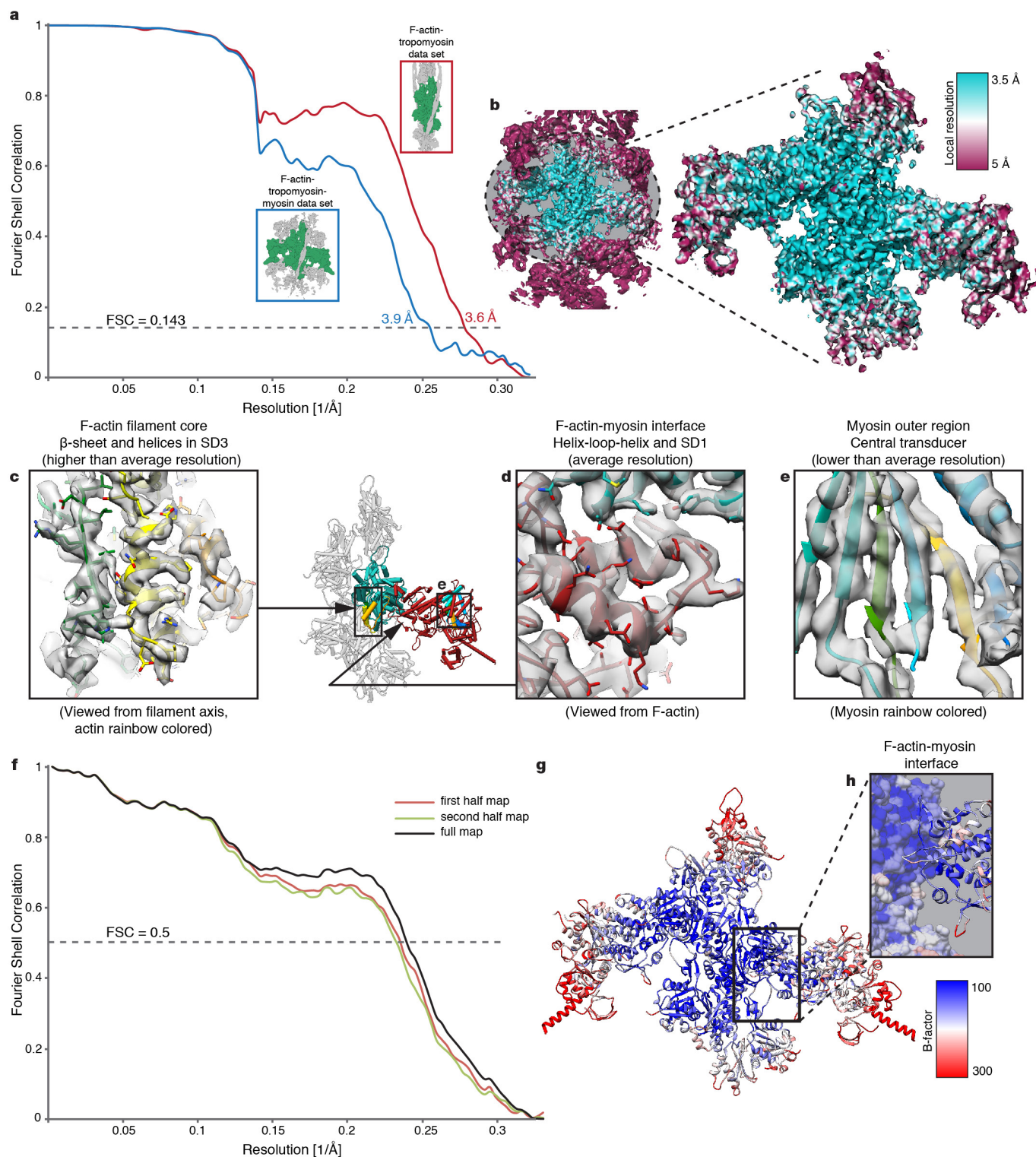
31. Müller, M. *et al.* Functional characterization of the human α -cardiac actin mutations Y166C and M305L involved in hypertrophic cardiomyopathy. *Cell. Mol. Life Sci.* **69**, 3457–3479 (2012).
32. Ohki, T., Ohno, C., Oyama, K., Mikhailenko, S. V. & Ishiwata, S. Purification of cytoplasmic actin by affinity chromatography using the C-terminal half of gelsolin. *Biochem. Biophys. Res. Commun.* **383**, 146–150 (2009).
33. Coulton, A., Lehrer, S. S. & Geeves, M. A. Functional homodimers and heterodimers of recombinant smooth muscle tropomyosin. *Biochemistry* **45**, 12853–12858 (2006).
34. Heissler, S. M. & Manstein, D. J. Comparative kinetic and functional characterization of the motor domains of human nonmuscle myosin-2C isoforms. *J. Biol. Chem.* **286**, 21191–21202 (2011).
35. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
36. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
37. Hohn, M. *et al.* SPARX, a new environment for cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
38. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
39. Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
40. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
42. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
43. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
44. Münnich, S., Pathan-Chhatbar, S. & Manstein, D. J. Crystal structure of the rigor-like human non-muscle myosin-2 motor domain. *FEBS Lett.* **588**, 4754–4760 (2014).
45. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
46. López-Blanco, J. R. & Chacón, P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).
47. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
49. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
50. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
51. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
52. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537–W541 (2012).
53. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
54. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
55. Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* **133**, 1–9 (2014).
56. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
57. Dausse, E. *et al.* Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and identification of a hot spot for mutations in the beta-myosin heavy chain gene. *J. Clin. Invest.* **92**, 2807–2813 (1993).
58. Richard, P. *et al.* Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* **107**, 2227–2232 (2003).
59. Yu, B. *et al.* Denaturing high performance liquid chromatography: high throughput mutation screening in familial hypertrophic cardiomyopathy and SNP genotyping in motor neurone disease. *J. Clin. Pathol.* **58**, 479–485 (2005).
60. Epstein, N. D., Cohn, G. M., Cyran, F. & Fananapazir, L. Differences in clinical expression of hypertrophic cardiomyopathy associated with two distinct mutations in the β -myosin heavy chain gene. A 908^{Leu→Val} mutation and a 403^{Arg→Gln} mutation. *Circulation* **86**, 345–352 (1992).
61. Geisterfer-Lowrance, A. A. *et al.* A molecular basis for familial hypertrophic cardiomyopathy: a β cardiac myosin heavy chain gene missense mutation. *Cell* **62**, 999–1006 (1990).
62. Blanchard, E., Seidman, C., Seidman, J. G., LeWinter, M. & Maughan, D. Altered crossbridge kinetics in the α MHC403/+ mouse model of familial hypertrophic cardiomyopathy. *Circ. Res.* **84**, 475–483 (1999).
63. Cuda, G., Fananapazir, L., Zhu, W. S., Sellers, J. R. & Epstein, N. D. Skeletal muscle expression and abnormal function of beta-myosin in hypertrophic cardiomyopathy. *J. Clin. Invest.* **91**, 2861–2865 (1993).
64. Woo, A. *et al.* Mutations of the beta-myosin heavy chain gene in hypertrophic cardiomyopathy: critical functional sites determine prognosis. *Heart* **89**, 1179–1185 (2003).
65. Van Driest, S. L. *et al.* Comprehensive analysis of the beta-myosin heavy chain gene in 389 unrelated patients with hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol.* **44**, 602–610 (2004).
66. Mohiddin, S. A. *et al.* Utility of genetic screening in hypertrophic cardiomyopathy: prevalence and significance of novel and double (homozygous and heterozygous) β -myosin mutations. *Genet. Test.* **7**, 21–27 (2003).
67. Marian, A. J. *et al.* A transgenic rabbit model for human hypertrophic cardiomyopathy. *J. Clin. Invest.* **104**, 1683–1692 (1999).
68. Lankford, E. B., Epstein, N. D., Fananapazir, L. & Sweeney, H. L. Abnormal contractile properties of muscle fibers expressing beta-myosin heavy chain gene mutations in patients with hypertrophic cardiomyopathy. *J. Clin. Invest.* **95**, 1409–1414 (1995).
69. Perrot, A. *et al.* Prevalence of cardiac beta-myosin heavy chain gene mutations in patients with hypertrophic cardiomyopathy. *J. Mol. Med.* **83**, 468–477 (2005).

70. Erdmann, J. *et al.* Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. *Clin. Genet.* **64**, 339–349 (2003).
71. Moolman, J. C., Brink, P. A. & Corfield, V. A. Identification of a new missense mutation at Arg403, a CpG mutation hotspot, in exon 13 of the β -myosin heavy chain gene in hypertrophic cardiomyopathy. *Hum. Mol. Genet.* **2**, 1731–1732 (1993).
72. Moolman-Smook, J. C., De Lange, W. J., Bruwer, E. C., Brink, P. A. & Corfield, V. A. The origins of hypertrophic cardiomyopathy-causing mutations in two South African subpopulations: a unique profile of both independent and founder events. *Am. J. Hum. Genet.* **65**, 1308–1320 (1999).
73. Greber-Platzer, S. *et al.* Beta-myosin heavy chain gene mutations and hypertrophic cardiomyopathy in Austrian children. *J. Mol. Cell. Cardiol.* **33**, 141–148 (2001).
74. Villard, E. *et al.* Mutation screening in dilated cardiomyopathy: prominent role of the beta myosin heavy chain gene. *Eur. Heart J.* **26**, 794–803 (2005).
75. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
76. Toydemir, R. M. *et al.* Mutations in embryonic myosin heavy chain (*MYH3*) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome. *Nature Genet.* **38**, 561–565 (2006).
77. Kamisago, M. *et al.* Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N. Engl. J. Med.* **343**, 1688–1696 (2000).
78. Daehmlow, S. *et al.* Novel mutations in sarcomeric protein genes in dilated cardiomyopathy. *Biochem. Biophys. Res. Commun.* **298**, 116–120 (2002).



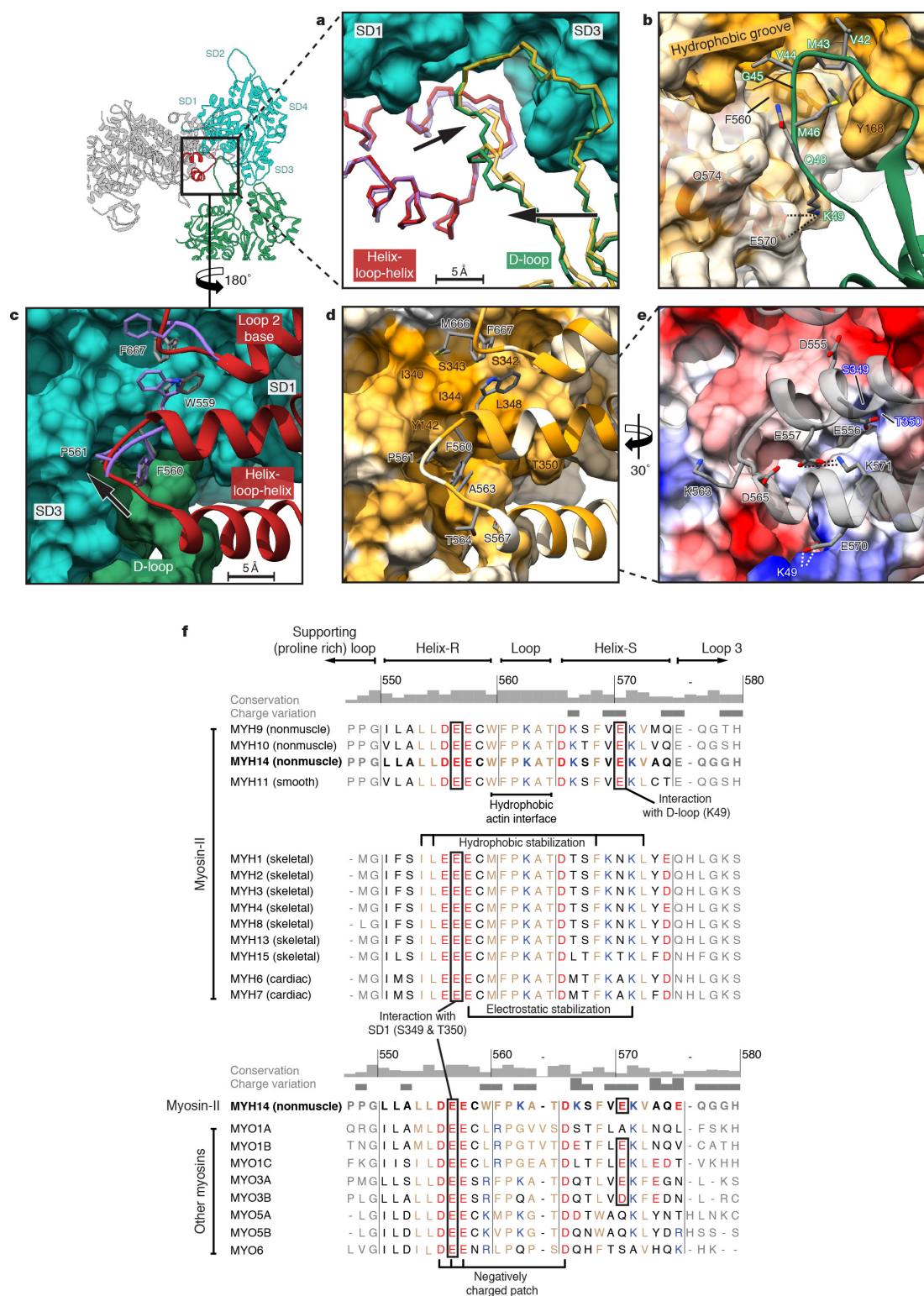
Extended Data Figure 1 | Micrographs, two-dimensional classifications and three-dimensional refinement. a–c, Representative of ~2,000 digital micrographs (a) and of 200 two-dimensional class averages of the F-actin–tropomyosin–myosin data set before (b) and after (c) three-dimensional refinement, respectively. Lower part of the micrograph is band-pass filtered to allow a better visualization of the filaments. Only filaments in rectangular boxes were chosen for refinement and bundled filaments were sorted out. d, e, Box dimension and angular distribution during three-dimensional refinement in side (d) and top (e) views. Histogram (few in blue to many in red) shows distribution

of projection direction of each boxed segment relative to the three-dimensional reconstruction (grey). f, Example of two class averages out of c that show secondary structure elements. g, Fit of F-actin–myosin model to assign characteristic domains of myosin (see coloured circles and boxes). h, i, Representative of 200 class averages of the reprocessed F-actin–tropomyosin data set before (h) and after (i) three-dimensional refinement. j, k, In class averages, secondary structure elements in F-actin (green) and the coiled-coil structure of tropomyosin (yellow) are visible. Scale bars in micrograph and class averages are 50 nm and 10 nm, respectively.



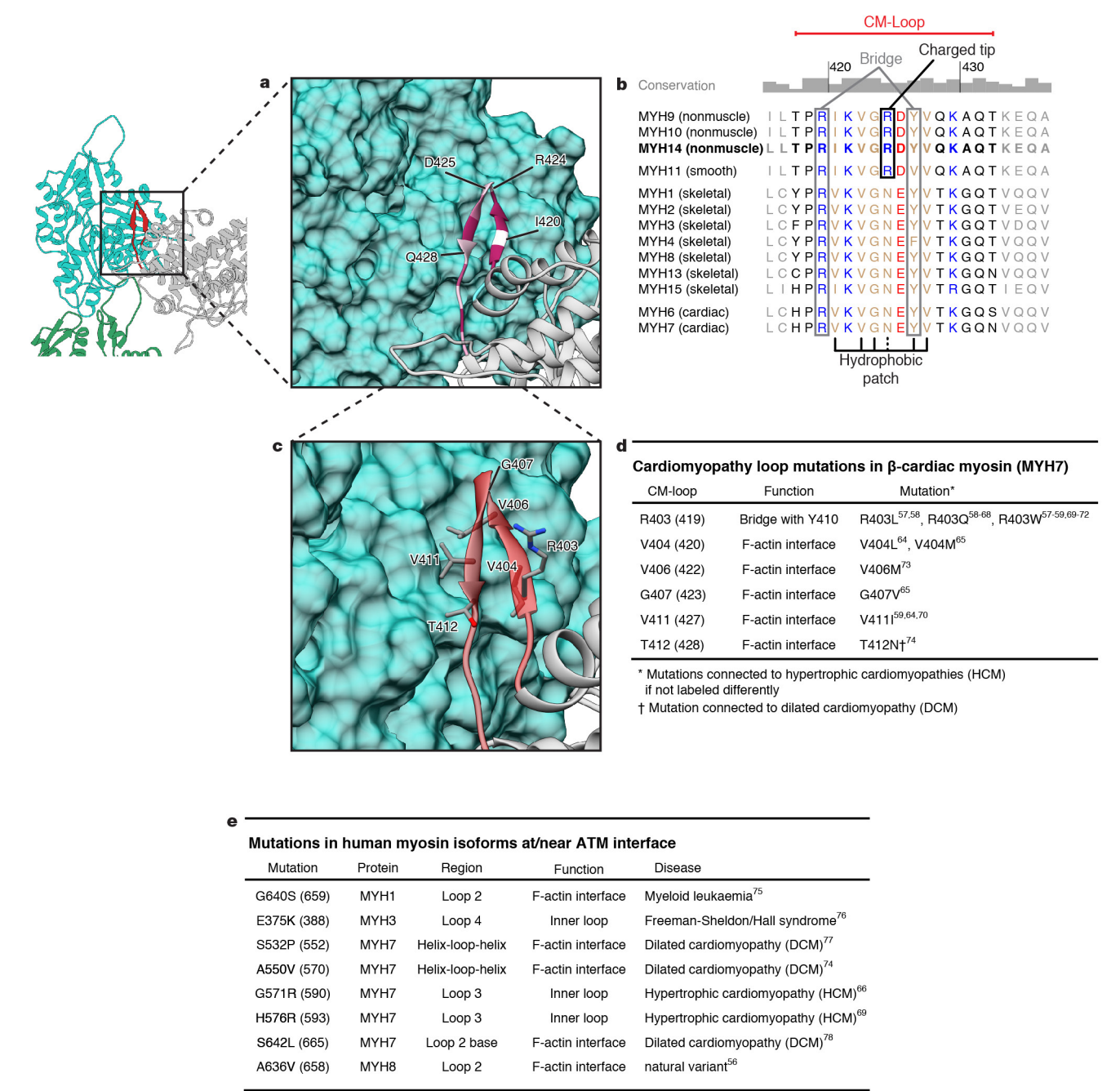
Extended Data Figure 2 | Resolution and model refinement of the actomyosin complex. **a**, FSC curves of the cryo-EM reconstruction of the F-actin–tropomyosin–myosin data set (blue) and the reprocessed data set of F-actin–tropomyosin (red). The average resolution ($\text{FSC}_{0.143}$) of the final electron density maps (central parts, green in subfigures) is estimated at 3.9 Å and 3.6 Å, respectively. Next subfigures illustrate only the actomyosin data set. **b**, Colour-coded local resolution of the full map and only finally refined part of the map (see Methods) estimated by ResMap⁴³. **c–e**, Representative regions with higher than the average resolution in the

F-actin filament core (**c**), the average resolution at the interface (**d**) and lower resolution in outer myosin parts (**e**). **f**, FSC curves of the model to each half map to check for overfitting, when the model was only refined versus the first half map. Black curve shows FSC between refined model and full map, when the model was refined against the full map (see Methods). **g**, **h**, B-factor distribution of final model from low (blue) to high (red) values. The absolute value strongly depends on the sharpening factor of the map, while the distribution shows the same gradient as the local resolution in **b**.



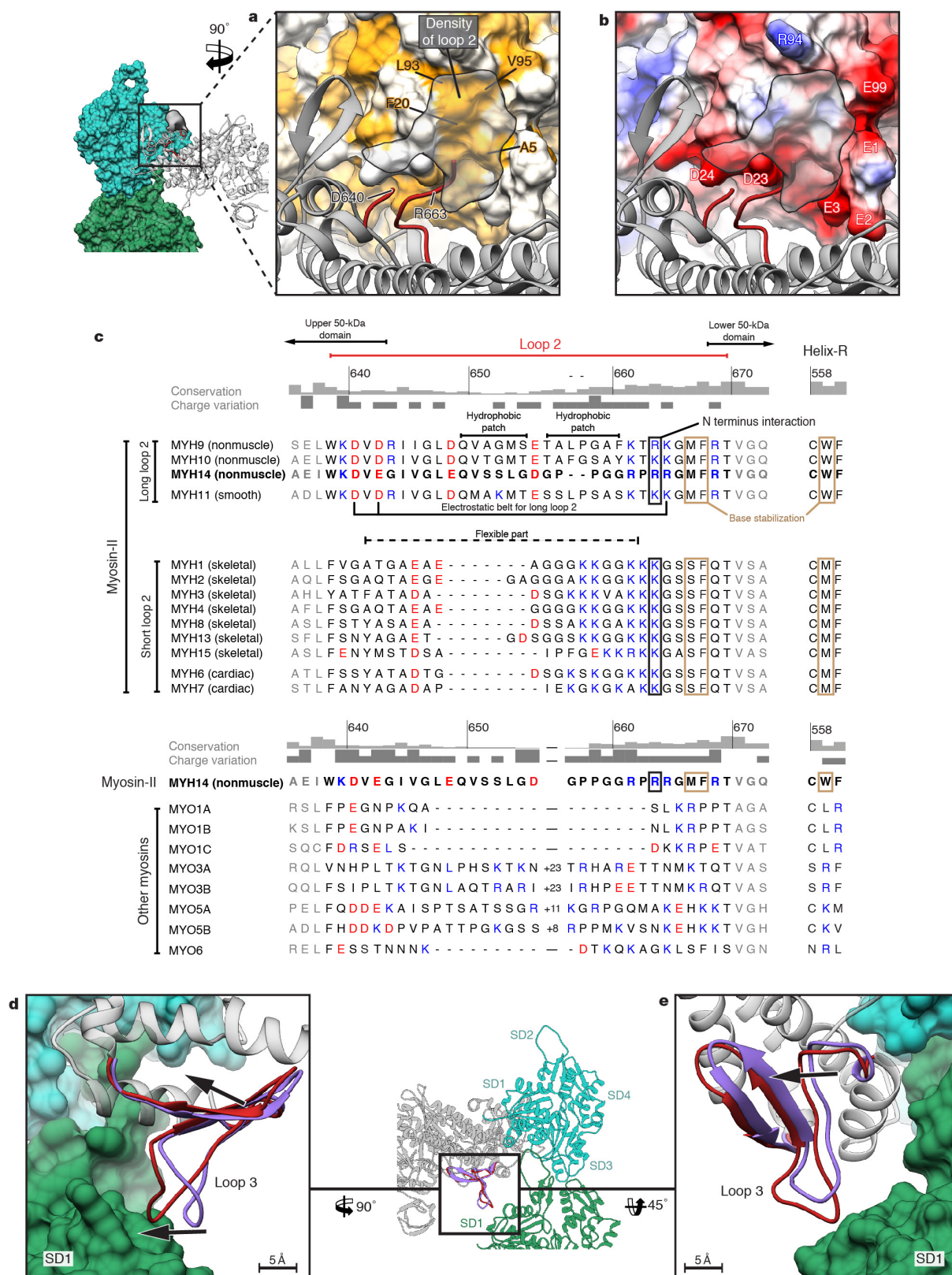
Extended Data Figure 3 | HLH motif bound to F-actin. **a**, Front view of F-actin and the HLH motif of the L50 domain of myosin show only small changes in loop regions while helices do not alter between weak (PPS state in purple, PDB accession number 514E) and strong binding (rigor state in red). The D-loop is moved towards the binding interface and is stabilized (A-state in yellow, M-state in green). Arrows indicate changes and scale bar is given. **b**, Same view as before shows the interface of myosin and F-actin in the rigor state. One possible salt bridge is highlighted with dotted lines. Surfaces are coloured from low (white) to high (yellow) hydrophobicity. **c–e**, Back view of the HLH motif and the base of loop 2 bound to central (SD1, SD3) and adjacent (D-loop) F-actin subunits. Comparison of rigor (red) and PPS state (purple, PDB accession number 514E) shows main differences (**c**). Final interaction of fully bound myosin

is given in **d**, **e**. Possible electrostatic interactions are indicated by dotted lines. F-actin surface is depicted per subunit colour (**c**), by hydrophobicity (**d**) or electrostatic Coulomb potential (**e**, $-10 \text{ kcal mol}^{-1}$ in red to $+10 \text{ kcal mol}^{-1}$ in blue). In all subfigures, coloured residue labels belong to F-actin. **f**, Sequence alignment of myosin (*H. sapiens* myosin-II, -I, -III, -V, -VI) in the region of the HLH (helix-R-loop-helix-S) motif. Important functions at the F-actin–myosin interface and roles in stabilizing these regions themselves are highlighted and labelled. Residue numbering refers to our published structure belonging to the sequence of NM-2C (depicted in bold type). Tissue localization of myosin-II is written in parentheses. We refer to the different myosin isoforms according to the nomenclature for the genes encoding the respective myosin heavy chains.



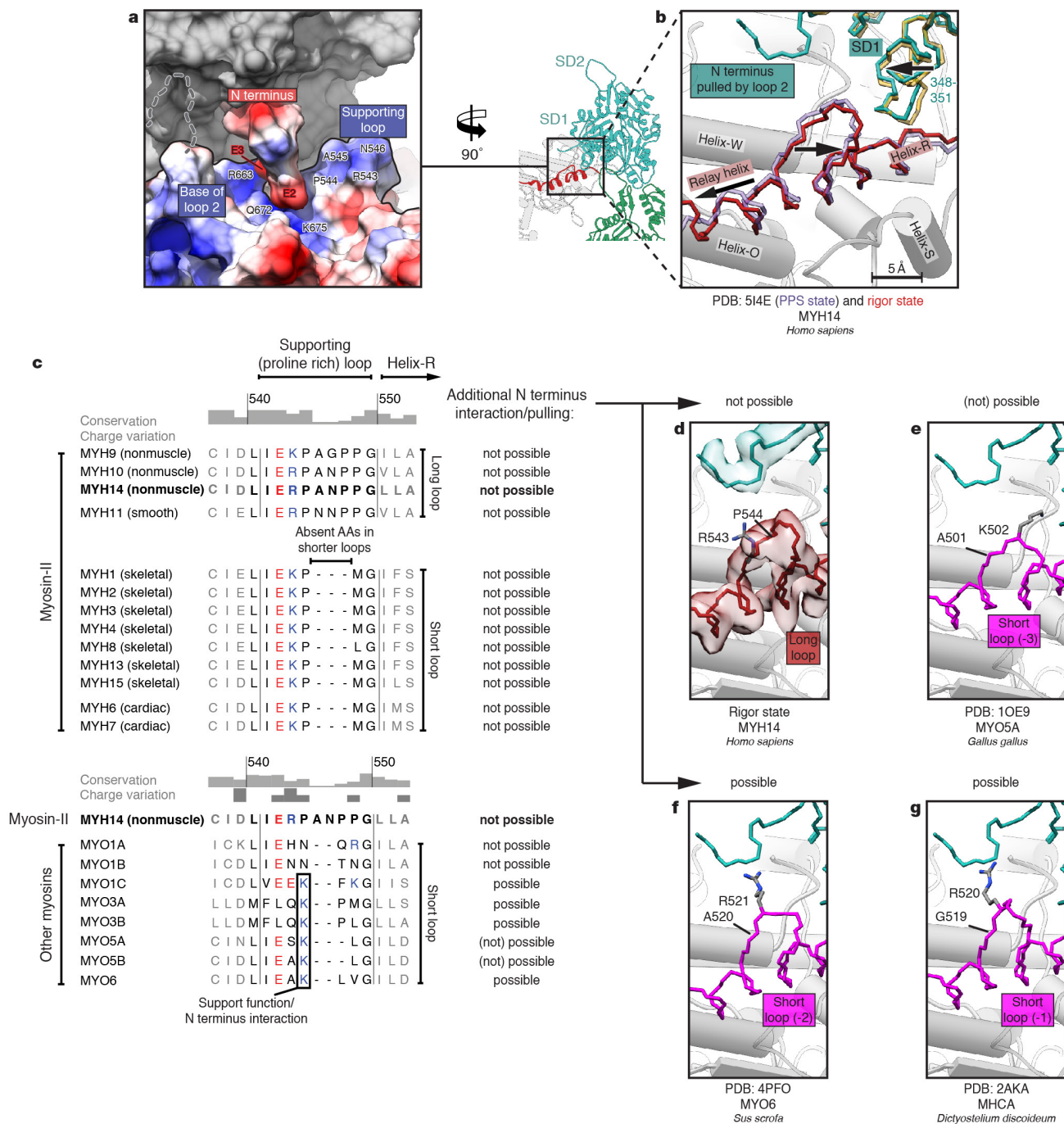
Extended Data Figure 4 | Cardiomyopathy loop and disease-causing mutations. **a**, Conservation of the CM-loop in the human myosin-II class is visualized as a model on F-actin (cyan) from low (white) to high (purple) conservation. **b**, Sequence alignment of the CM-loop region of the human myosin-II class. Important functions at the F-actin–myosin interface are highlighted and labelled. Residue numbering refers to our published structure belonging to the sequence of NM-2C (depicted in bold type). Tissue localization of myosin-II is written in parentheses. We refer

to the different myosin isoforms according to the nomenclature for the genes encoding the respective myosin heavy chains. **c**, **d**, Mutations in β -cardiac myosin (MYH7) can lead to cardiomyopathies. Corresponding residues in β -cardiac myosin are illustrated with our rigor state model (**c**) and known mutations^{57–74} are listed (**d**). **e**, Table of known disease-causing mutations at the actomyosin interface^{56,66,69,74–78}. Numbers in parentheses give respective residue position in our published structure of NM-2C. Localization is described in parentheses.



Extended Data Figure 5 | Loop 2 and loop 3 on F-actin. **a, b**, Density map (grey) corresponding to the flexible part (residues 641–661) of loop 2 (red). F-actin is shown as surface model coloured from low (white) to high (yellow) hydrophobicity (**a**) or electrostatic Coulomb potential (**b**, $-10 \text{ kcal mol}^{-1}$ in red to $+10 \text{ kcal mol}^{-1}$ in blue). Residue labels belonging to F-actin are coloured as surface colours. **c**, Sequence alignment of myosin (*H. sapiens* myosin-II, -I, -III, -V, -VI) in the region of loop 2 and helix-R of the HLH region. Important functions at the F-actin–myosin interface and in stabilizing these regions are

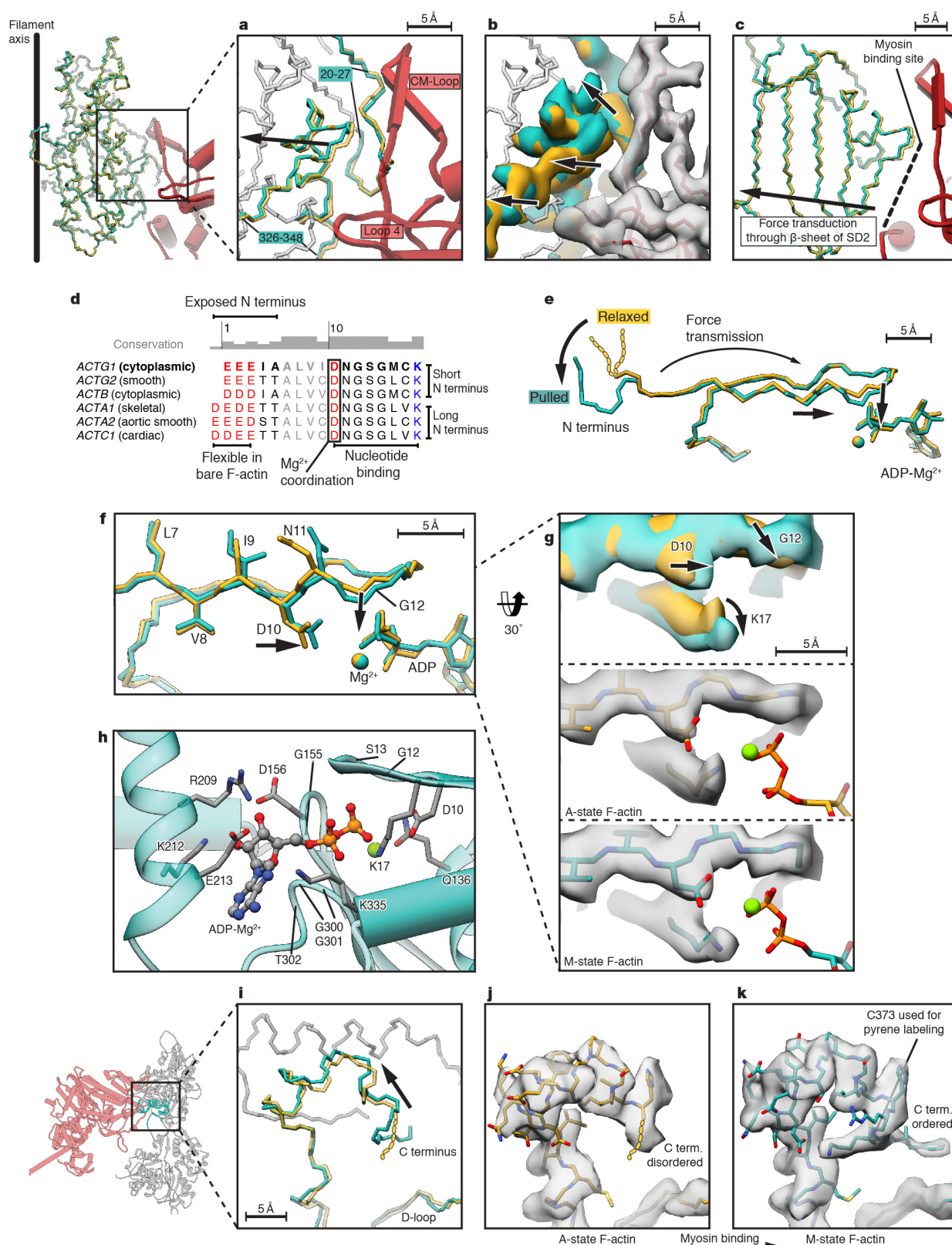
highlighted and labelled. Residue numbering refers to our published structure belonging to the sequence of NM-2C (depicted in bold). Tissue localization of myosin-II is written in parentheses. We refer to the different myosin isoforms according to the nomenclature for the genes encoding the respective myosin heavy chains. **d, e**, Changes between rigor (red) and PPS state (purple) in the loop 3 region relative to the rest of lower 50-kDa domain when bound to F-actin. Movements are indicated by black arrows and scale bars are given.



Extended Data Figure 6 | Sequence-dependent interaction of supporting loop with the N terminus of F-actin.

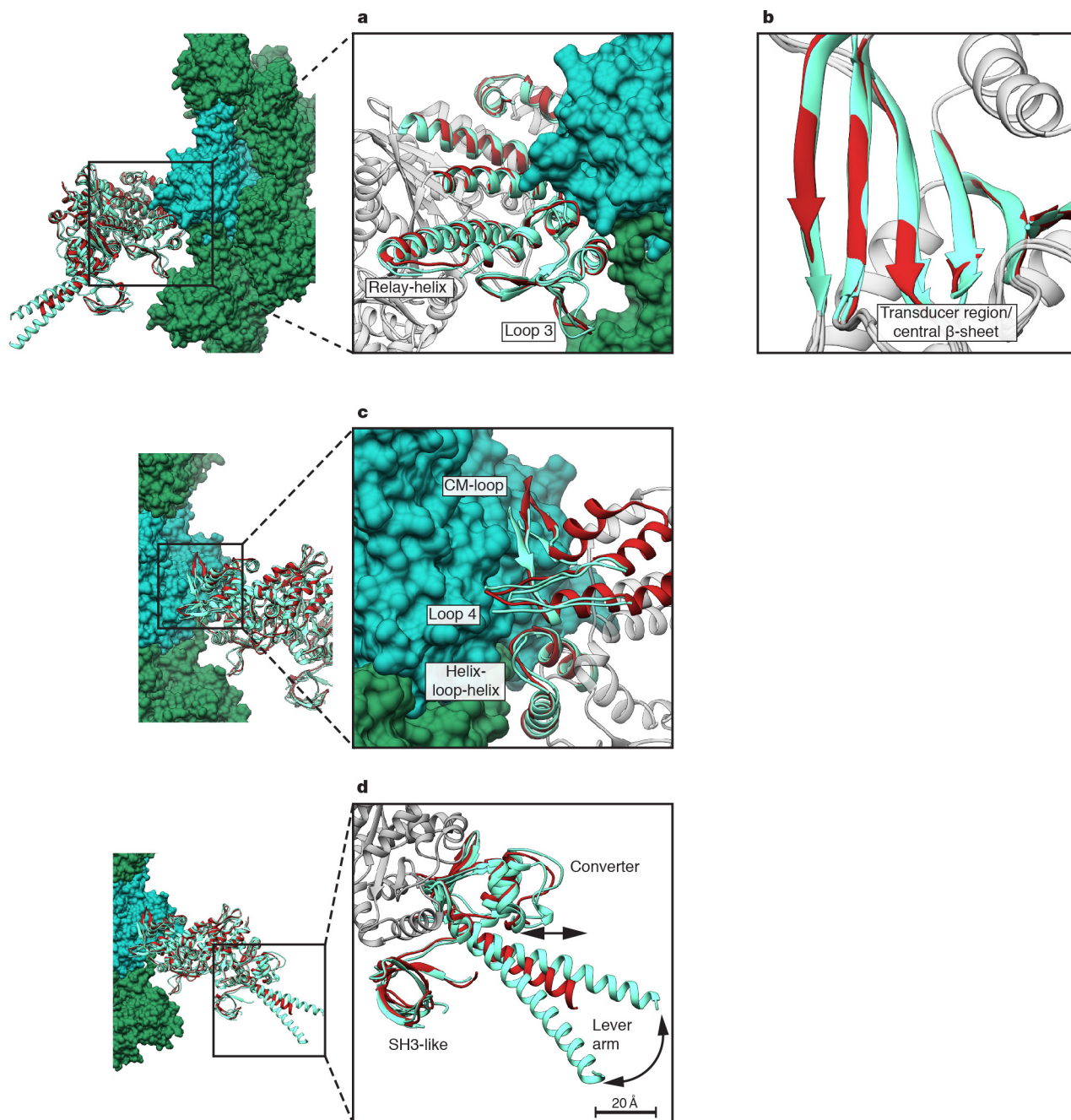
a, Surface of myosin and N terminus is depicted by electrostatic Coulomb potential ($-10 \text{ kcal mol}^{-1}$ in red to $+10 \text{ kcal mol}^{-1}$ in blue). Involved charged residues are labelled. **b**, Position of the proline-rich loop (supporting loop) located between relay helix and helix-R slightly differs between the PPS (purple, PDB accession number 5I4E) and rigor state (red) and shows no direct interaction with the N terminus of F-actin. Regions at the surface of SD1 are pulled to the actomyosin interface indicated by an arrow and a scale bar is given (F-actin in A-state is depicted in yellow; F-actin in M-state is depicted in cyan). **c**, Sequence alignment of myosin (*H. sapiens* myosin-II, -I, -III, -V, -VI) in the region of the supporting loop. Different lengths of the loop and a possible supporting function are given in the last column.

Residue numbering refers to our published structure belonging to the sequence of NM-2C (depicted in bold). Tissue localization of myosin-II is written in parentheses. We refer to the different myosin isoforms according to the nomenclature for the genes encoding the respective myosin heavy chains. **d–g**, Comparison of prominent properties in the supporting loop of different myosin classes (comparative models in purple) and their ability to undergo a direct interaction with the N terminus. Main differences are length of loop (numbers give absent amino acids relative to long loop) and position of the prominent positive-charged amino acid (R or K). Only an arginine or lysine sitting on the top would allow a direct interaction (**e–g**), while a sideward-oriented arginine (**d**) or a short loop (**e**) disables or reduces a possible interaction, respectively. In addition, respective densities (**d**) of the cryo-EM map are displayed.



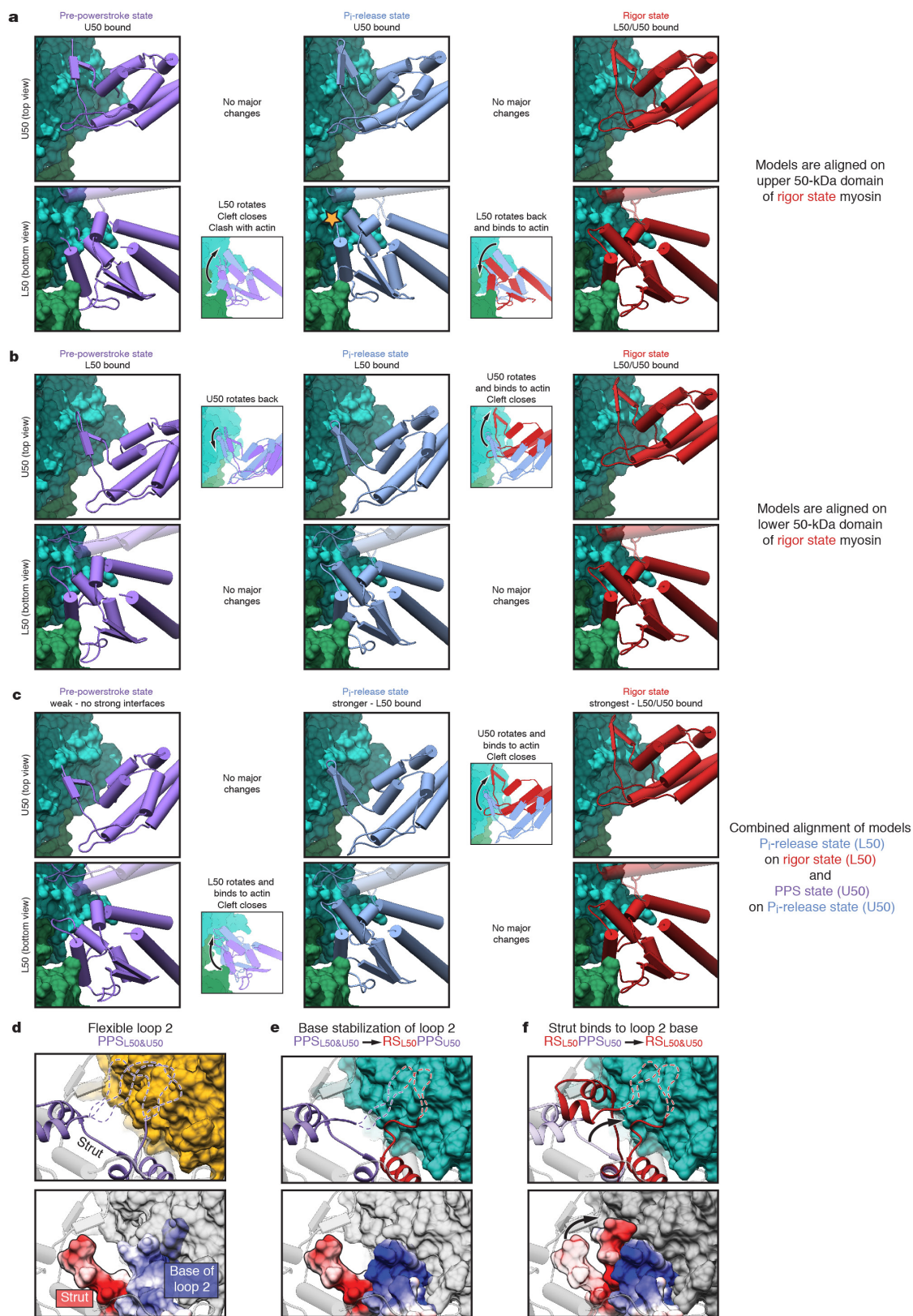
Extended Data Figure 7 | Myosin-induced conformational changes in F-actin. **a–c**, Comparison of bare F-actin (A-state, yellow) with myosin-bound F-actin (M-state, cyan). Myosin is depicted in red. Either models (**a**) or representative parts of the electron density maps (**b**) illustrate conformational changes in F-actin (**c**). **d**, Sequence alignment of the N terminus region of human actin isoforms. Residue numbering refers to our published structure belonging to the sequence of non-muscular γ 1-actin (*ACTG1*, depicted in bold type). To prevent confusion, the gene names instead of protein names are given. Localization is written

in parentheses. **e**, N terminus and nucleotide binding region of F-actin undergo small changes (highlighted with arrows) through transmitted force of N terminus pulling. **f**, **g**, Close-ups of structural changes at the nucleotide binding site. **h**, Coordination of ADP and Mg^{2+} in the nucleotide binding cleft in M-state F-actin. **i–k**, Myosin binding induces a stabilization and shifting of the C terminus towards SD1 of F-actin. C373, which was used for pyrene labelling of F-actin, is part of the C-terminal region. Scale bars are given in the subfigures.



Extended Data Figure 8 | Comparison of rigor and rigor-like myosin structures. **a–d**, Close-ups of superimposed models from rigor-like structures (nucleotide-free myosin crystal structures in light green, PDB accession number 4PD3 (ref. 44) and 1OE9 (ref. 27)) with our rigor cryo-EM structure (red). F-actin is shown as surface model (green, cyan). Illustrated domains are labelled and coloured, while the rest of myosin is shown in grey from the rigor state model. Most regions do not show

conformational differences (**a**, **b**), but the surface loops of myosin (CM-loop, loop 3 and loop 4) interacting with F-actin differ slightly in the rigor from the rigor-like structures (**a**, **c**). In contrast to the cryo-EM structure, loops at the interface (**a**, **c**) between F-actin and myosin are not always resolved in crystal structures. Major structural differences in the lever arm and converter regions are indicated by arrows and a scale bar is given (**d**).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Different alignments of models for weak to strong binding of myosin and strut attraction to the base of loop2 promotes cleft closure. **a–c**, Three possible alignments of myosin in the PPS (first column, purple, PDB accession number 5I4E), P_i-release (second column, blue) and rigor (third column, red) states are illustrated with respect to F-actin. For better visualization, differences in F-actin are not shown and F-actin is only depicted in the M-state (green, cyan). The P_i-release state represents a homology model of NM-2C based on a crystal structure of myosin in the P_i-release state⁷ (PDB accession number 4PFO, see Methods). All models are either aligned to the U50 domain (**a**) or the L50 domain (**b**) of the rigor state. In **c**, the model of the P_i-release state was first aligned to the L50 domain of the rigor state. The PPS state was then aligned to the U50 domain of model of the P_i-release state. The first row in each subfigure shows changes in the U50 domain from the top (for a better visualization L50 was deleted). The second row shows

the L50 domain from the bottom (U50 is transparent). Possible clashes are indicated by a yellow star (**a**). **d–f**, Binding mechanism of the strut, connecting L50 and U50 domains, to the stabilized base of loop 2. To illustrate the conformational changes, the respective regions in the PPS state (PPS, purple, PDB accession number 5I4E) and rigor state (RS, red) of myosin have been partly overlaid. The rest of myosin is shown in grey. L50 binds to F-actin (A-state, yellow) (**d**, **e**). The base of loop 2 is stabilized by F-actin (**e**) and attracts the negatively charged strut with its positive patch. This promotes the binding of the strut, shifting the equilibrium to a closed conformational state of myosin (**f**). Flexible parts of loop 2 are indicated as dotted lines. Lower panels show surfaces of the same regions as in the upper panels coloured by electrostatic Coulomb potential. For better visualization, the upper parts of the strut were removed. Surface of F-actin is depicted in transparent grey.

Extended Data Table 1 | Data collection and refinement statistics

Data collection			
	Magnification	x122,270	
	Defocus range (μm)	0.7-2.8	0.8-2.6
	Voltage (kV)	300	
	Microscope	Titan Krios	
	Camera	Falcon 2	
	Frame recording time (s)	0.085-0.475	
	Number of frames	7	
	Electron dose ($\text{e}/\text{\AA}^2$)	16	
	Pixel size (\AA)	1.1	
Particle statistics			
	Box size (px)	256	
	Boxing distance (px)	29	
	Rise* (\AA)	27.5	
	Azimuthal rotation* ($^\circ$)	166.9	
	Particles	118,000	91,000
Model composition			
		F-actin-myosin	Bare F-actin
	Non-hydrogen atoms	26,477	14,450
	Protein residues	3,354	1,845
	Ligand (ADP/Mg ²⁺)	135/5	135/5
Refinement			
	Resolution (\AA)	3.9	3.6
	Map sharpening <i>b</i> factor (\AA^2)	-200	-200
	Average B-factor (\AA^2)	180	98
	R factor	0.34	0.33
	Fourier Shell Correlation	0.84	0.83
Rms deviations			
	Bonds (\AA)	0.015	0.013
	Angles ($^\circ$)	1.83	1.74
Validation			
	Molprobity score	2.24	1.82
	Clashscore, all atoms	14.16	8.93
	Poor rotamers (%)	1.57	1.36
Ramachandran plot			
	Favored (%)	93.23	96.38
	Allowed (%)	5.14	3.07
	Outliers (%)	1.63	0.55

*Helical symmetry parameters were estimated after C1 refinement (see Methods for further details)

Refinement statistics are given after the last step of refinements of the actomyosin and bare F-actin data set. Rms, root mean square.

REVISION THEORY

Time to take stock.

BY BLAIZE M. KAYE

Themba reread what he'd written. "I must know if this will work. It's 11:35 p.m., 12 Sept. 2015. I'm in the garage. In 2 minutes, I'm going to open the third drawer down on the left side of my workbench."

That was it.

There were two possibilities.

One: the drawer would remain as empty as it had been since the day he'd bought the bench. That wouldn't necessarily mean that the device didn't work, but it would be disappointing.

Two: the drawer would no longer be empty.

He folded the note carefully and focused on what he intended to do with it. He would place it in a white envelope, address it to himself, and deposit it in the red and blue postbox next to the notice board at the supermarket.

He waited.

11:37:59 p.m.

Themba opened the drawer. Inside lay a yellowed envelope.

He trembled as he retrieved the letter and opened it. There, written in a faded blue script he didn't recognize, was a simple response:

"Yes, it works."

Themba sat, stunned for a moment. Then, a wave of relief passed over him. Graduate school, the disastrous postdoc, the humiliation of taking a job in industry while his peers were getting their own labs. None of that meant anything now. Not next to this.

"It works," he said. And then, correcting himself: "It's *going* to work."

He glanced at the drawer and turned his attention back to the letter. He'd seen it before, just after they'd delivered the workbench. He remembered opening the drawer a hundred times since and seeing the letter just lying there. Why hadn't he read it, he wondered, or just thrown it away?

This was to be expected when reweaving history, he thought. Every moment from the intervention to the present is affected — and the whole course of the future as well. These weren't false memories, he'd *really* seen the letter in the drawer before and yet, in another sense, he *couldn't* have.

He picked up his own note and added:

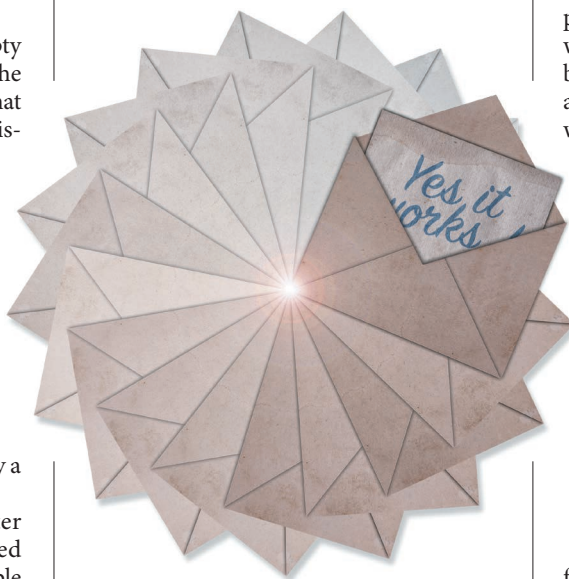
"PS I've been struggling with counterfactual dampening, don't tell me how to do

it, but am I on the right track? Is it even a problem?"

The letter in the drawer had always read:

"I don't know anything about 'counterfactual dampening' but, yes, your machine works."

Had the letter changed? He remembered adding the PS just a moment ago, but he was



certain that the response had always mentioned the dampening issue. And who could be using the machine without even recognizing one of the most fundamental terms in temporal and possible world mechanics?

He added to his note:

"PPS Who is this?"

And the response from the drawer had always read:

"Dear Dad, this is Pat. I don't know anything about 'counterfactual dampening', but, yes, your machine works."

Pat. Patricia? His little girl, whom he'd bathed and changed just hours earlier. Whom he'd kissed goodnight before sneaking into his workshop.

Thema again added to his note.

"PPPS Pat. Do I *know* you're writing to me?"

And Patricia's response had always read:

"Dad, you don't know that I'm doing this, but I wanted to tell you that, yes, it works. I don't know anything about 'counterfactual dampening', sorry. I love you and miss you. Patricia."

Themba stared at the note for a long while. Something about the tone bothered him.

He crossed out the line he'd just added and wrote:

"PPPS *Why* don't I know that you're writing this to me?"

And Patricia's response had always read:

"Dear Dad, I miss you so much. I suppose you must have been on the right track with the 'counterfactual dampening' issue because, yes, your machine works. I've spent a long time trying to decide how, and even whether, I should answer the last question

in your note. We missed you so much during those early years, but Mom said your work was important, that it could change everything. Once you'd shown that you could move things through time, we thought we'd have you back. But it just got worse. You said there were still problems with the device, and we saw less of you then. Mom and I left. Miriam found you. It was a heart attack. Your note was among your papers. I haven't told Mom about it. It would make her too sad.

I love you.

Patricia."

He'd been crying from the moment he'd first opened and read Patricia's letter. There was no relief or joy in this success, not if it cost him his family and, ultimately, his life.

He disconnected the device and placed it in the open drawer. He was done with it.

Then he picked up the letter he'd written, the letter that would eventually end up in Patricia's hands, and tore it to pieces while focusing on his intention never to mail it.

He wanted to hold his daughter and his wife.

He wanted to meddle with time no longer.

He threw the scraps of paper into the small red wastebasket next to the bench.

He reached for Patricia's response but it was no longer there — it had never been there — and Themba was left with his hand suspended in mid-air with the strongest feeling that he was about to do something important but had no idea what it was.

He supposed it had something to do with his work and so he opened the drawer, retrieved the device, and began running diagnostics. There was still much to do. ■

Blaize M. Kaye lives in Kwazulu-Natal, South Africa, with his wife and child. He spends his time thinking about books, brains and computers.

ILLUSTRATION BY JACEY